МАЛАЯ МАТЕМАТИЧЕСКАЯ ЭНЦИКЛОПЕДИЯ

Э. ФРИД, И. ПАСТОР, И. РЕЙМАН,
П. РЕВЕС, И. РУЖА

1828

AKADÉMIAI KIADÓ
ИЗДАТЕЛЬСТВО АКАДЕМИИ НАУК ВЕНГРИИ
БУДАПЕШТ 1976

Данные оригинала

FRIED E. — PÁSZTOR I. — REIMAN I. — RÉVÉSZ P. — RUZSA I.
MATEMATIKAI KISENCIKLOPÉDIA

Szerkesztők:

LUKÁCS ERNŐNÉ és TARJÁN REZSŐNÉ

Gondolat Könyvkiadó, Budapest

Перевели с венгерского

Я. КОЧИШ и М. СОКОЛОВ

Редакторы перевода

Ю. Н. БЛАГОВЕЩЕНСКИЙ и Б. КОЧИШ

ISBN 963 05 0844 3

СОДЕРЖАНИЕ

АЛГЕБРА

ПРЕДМЕТ АЛГЕБРЫ

Первоначально алгебра была разделом математики, занимавшимся решением уравнений. В отличие от геометрии, аксиоматического построения алгебры не существовало до середины XIX века, когда появился принципиально новый взгляд на предмет и характер алгебры. Исследования стали все больше направляться на изучение так называемых алгебраических структур. Это имело два преимущества. С одной стороны, были уточнены области, для которых справедливы отдельные теоремы, с другой стороны, появилась возможность использовать одни и те же доказательства в совершенно разных областях. Такое разделение алгебры просуществовало до середины XX века и нашло свое выражение в том, что появились два названия: «классическая алгебра» и «современная алгебра». Последнюю больше характеризует другое название: «абстрактная алгебра». Дело в том, что для этого раздела — впервые в математике — была характерна полная абстракция. При доказательстве отдельных теорем абстрактной алгебры не принимаются во внимание даже математические свойства объектов, точнее, рассматриваются лишь те формальные свойства, которые необходимы для доказательства теоремы. Эти свойства принимаются затем за аксиомы. Именно поэтому получаемые теоремы оказываются справедливыми во всякой системе, в которой выполняются принятые аксиомы. С середины XX века раздел, именовавшийся современной алгеброй, стал называться абстрактной алгеброй или просто «алгеброй». Дело в том, что в это время стали появляться новые направления, где рассматривались не отдельные структуры, а типы структур; эта новая ветвь получила название современной алгебры.

ПОСТРОЕНИЕ ПОНЯТИЯ ЧИСЛА

НАТУРАЛЬНЫЕ ЧИСЛА

Представления о числах пришли из действительности. В математике из этих представлений стараются выбрать наиболее существенные, их принимают за *аксиомы*, а остальное доказывают логическим путем.

Понятие числа родилось в результате абстрагирования. Сначала возникло понятие натуральных чисел. Это числа 1, 2, 3 и так далее, образующие последовательность, которую можно неограниченно продолжать. Для возникновения других чисел потребовалось новое абстрагирование.

О натуральных числах достаточно предположить, что начав с 1, мы путем обычного счета будем каждый раз получать новые натуральные числа, и таким образом получим любое число ряда.

ПОЛНАЯ ИНДУКЦИЯ

Описать сформулированное выше свойство натуральных чисел с помощью аксиом оказалось бы значительно сложнее. Приведенная формулировка уже заключает в себе тот факт, что для натуральных чисел может быть использован так называемый метод *полной индукции*.

Индукцией называют метод доказательства, при котором на основе частных случаев делается заключение относительно общего случая. Например, многочисленные опыты показывают, что камень, опущенный с некоторой высоты, падает вниз. Отсюда мы делаем вывод, что опущенный камень *всегда* должен упасть. Этот метод используется в естественных науках, где им приходится довольствоваться, так как все возможные случаи испробовать нельзя. Однако в математике такая индукция не может быть принята в качестве метода доказательства. Подтвердим это примером. Рассмотрим многочлен $x^2 - x + 41$ (см. стр. 65). Подставляя вместо x значения 0, 1, 2, 3, 4, и 5, получаем, соответственно, следующие значения многочлена : 41, 41, 43, 47, 53 и 61. Нетрудно видеть, что все эти числа — простые. Используя индукцию, мы пришли бы к выводу, что значением этого многочлена всегда оказывается простое число. Проделаем еще несколько «опытов» :

значение x:	6	7	8	9	10	11	12	13	14	15	16
значение многочлена:	71	83	97	113	131	151	173	197	223	251	281

Все полученные значения опять оказались простыми числами. Этот результат еще раз показывает, что наше предположение правильно. Продолжая «опыты» дальше, мы получали бы все новые подтверждения правильности сделанного предположения. До значения $x = 40$ значение многочлена всякий раз оказывается простым числом. Если же $x = 41$, то $x^2 - x + 41 = 41^2 - 41 + 41 = 41^2$. Это число уже не является простым.

Следовательно, в математике встречаются случаи, когда некоторое утверждение (например, утверждение о том, что значение многочлена $x^2 - x + 41$ является простым числом) оказывается верным для достаточно большого количества натуральных чисел (в приведенном примере до $x = 40$), но для достаточно большого натурального числа (в нашем примере это «большое» число равнялось 41) уже неверно. По этой причине в математике индукцию нельзя использовать для доказательств.

Однако с помощью казалось бы незначительного изменения из индукции можно получить метод, применимый для математических доказательств. Этот метод называется *полной индукцией*. Рассмотрим пример на использование полной индукции.

Докажем, что сумма первых n нечетных чисел равна n^2. Начнем опять с «опытов». Если $n = 1$, то нужно рассмотреть первое нечетное число, то есть 1, и действительно, $1 = 1^2$. Если $n = 2$, то $1 + 3 = 4 = 2^2$. Для $n = 3$ имеем: $1 + 3 + 5 = 9 = 3^2$. Далее получаем: если $n = 4$, то $1 + 3 + 5 + 7 = 16 = 4^2$; если $n = 5$, то $1 + 3 + 5 + 7 + 9 = 25 = 5^2$; если $n = 6$, то $1 + 3 + 5 + 7 + 9 + 11 = 36 = 6^2$. Рассмотрим теперь случай $n = 7$, но вместо того, чтобы выполнять суммирование, попробуем вывести результат из предыдущего. Для этого рассмотрим первые шесть членов суммы $1 + 3 + 5 + 7 + 9 + 11 + 13$. В случае $n = 6$ мы видели, что $1 + 3 + 5 + 7 + 9 + 11 = 6^2$, следовательно, $1 + 3 + 5 + 7 + 9 + 11 + 13 = 6^2 + 13$. Отсюда с помощью простых преобразований получаем: $6^2 + 13 = 6^2 + 12 + 1 = 6^2 + 2 \cdot 6 + 1 = (6 + 1)^2 = 7^2$. В этом рассуждении мы никак не использовали того факта, что утверждение доказывается именно для $n = 7$. Предположим, что для первых k нечетных чисел мы уже знаем, что их сумма равна k^2. Тогда сумму первых $k + 1$ нечетных чисел можно получить, прибавив к сумме k нечетных чисел $(k + 1)$-ое нечетное число. Но $(k + 1)$-ое нечетное число есть $2k + 1$, а сумма первых k нечетных чисел по предположению равна k^2. Следовательно, сумма первых $k + 1$ нечетных чисел равна $k^2 + (2k + 1) = k^2 + 2k + 1 = (k + 1)^2$, что и требовалось доказать. Мы уже знаем, что сумма первых 7 нечетных чисел равна 7^2. Если в предыдущее доказательство вместо k подставить 7, то мы получим, что сумма первых 8 нечетных чисел равна 8^2. Зная это, подставим теперь вместо k число 8, тогда получим, что сумма первых 9 нечетных чисел равна 9^2 и так далее. Повторяя этот процесс шаг за шагом, можно убедиться, что для произвольного натурального числа n сумма первых n нечетных чисел равна n^2.

Это и есть метод полной индукции, который можно сформулировать следующим образом:

пусть имеется некоторое утверждение, касающееся натуральных чисел, о котором известно, что

1. утверждение справедило для 1.
2. Если утверждение справедливо для какого-либо натурального числа k, то оно справедливо и для числа $k + 1$.

Тогда это утверждение справедливо для любого натурального числа n.

Мы не сказали здесь, что понимается под словом «утверждение». Строгое определение можно дать лишь после основательного ознакомления с математической логикой. Здесь же мы условимся рассматривать лишь такие утверждения, которые могут быть сформулированы математически.

ДЕЙСТВИЯ НАД НАТУРАЛЬНЫМИ ЧИСЛАМИ

Действия над натуральными числами можно определить с помощью обычного процесса счета.

Сложение. Пусть a и b — натуральные числа.

Начнем считать с числа $a + 1$ и, отсчитав b раз, остановимся. Будем говорить, что в результате мы прибавили к натуральному числу a натуральное число b. Последнее число, полученное при счете, назовем суммой натуральных чисел a и b и обозначим через $a + b$.

При сложении используются следующие названия: a и b называются *слагаемыми*, а $a + b$ — *суммой*.

Перестановочное свойство. По смыслу не одно и то же, прибавим ли мы к натуральному числу a (путем счета) натуральное число b или наоборот, к натуральному числу b прибавим натуральное число a. Тем не менее, суммы $a + b$ и $b + a$ оказываются равными между собой, то есть $a + b = b + a$. Это свойство сложения называется *перестановочным или коммутативным*. В коммутативности сложения можно убедиться не только на примерах, выполнение этого свойства можно доказать с помощью полной индукции.

Ассоциативное свойство. Тем же путем можно доказать и *ассоциативность* сложения, согласно которому прибавляя к сумме двух натуральных чисел третье натуральное число, мы получим тот же результат, как если бы к первому числу прибавили сумму второго и третьего. В буквенных обозначениях:

$$\boxed{(a + b) + c = a + (b + c)\,.}$$

Группируя слагаемые произвольным образом (и в случае большего числа слагаемых), мы всякий раз будем получать одну и ту же сумму, поэтому при сложении скобки можно опустить. Предыдущую сумму, например, можно записать и просто в виде $a + b + c$, то есть

$$\boxed{a + b + c = (a + b) + c = a + (b + c)\,.}$$

Умножение. Если все слагаемые равны между собой, то сумму можно обозначить короче, записав друг за другом количество слагаемых и само слагаемое, и поставив между ними точку (или знак $\times$). Таким образом $b + b = 2 \cdot b$, $b + b + b = 3 \cdot b$ и так далее. В этом случае мы говорим об *умножении*. Другими словами, $a \times b = ab$ — это сумма, состоящая из a слагаемых, каждое из которых есть b.

Если нет суммы, а только одно число, то его можно рассматривать как сумму, состоящую из одного слагаемого. В этом случае $1 \cdot b = b$. При умножении a и b называются *сомножителями*, а $a \cdot b$ — *произведением*. Можно доказать, что умножение также обладает свойствами коммутативности и ассоциативности:

$$\boxed{\begin{array}{ll} a \cdot b = b \cdot a & \text{и} \\ (a \cdot b) \cdot c = a \cdot (b \cdot c)\,. & \end{array}}$$

Распределительное свойство. Кроме того, можно доказать, что имеет место *дистрибутивность* (или *распределительное свойство*) умножения относительно сложения, то есть для любых трех натуральных чисел a, b и c справедливо следующее соотношение:

$$\boxed{(a + b) \cdot c = (a \cdot c) + (b \cdot c)\,.}$$

Это соотношение связывает оба основных арифметических действия. *Знак умножения* — если это не приводит к недоразумению, — *можно опус-*

тить, то есть можно писать

$$\boxed{(a+b)c = (a \cdot c) + (b \cdot c) = (ac) + (bc)\,.}$$

Обычно принято считать, что умножение «связывает сильнее», чем сложение. Это означает, что при умножении следует заключать в скобки лишь те сомножители, которые содержат несколько членов (чисел). Так, например, в выражении $(a+b)c$ первый сомножитель заключается в скобки, а в выражении $(ac)+(bc)$ скобки можно опустить. Окончательно свойство дистрибутивности можно выразить следующим образом:

$$\boxed{(a+b)c = ac + bc \qquad \text{или} \qquad ac + bc = (a+b)c\,.}$$

Мы умышленно записали это тождество дважды, поменяв местами правую и левую части. Дело в том, что записанные тождества имеют разный смысл. Первое из них выражает правило раскрытия скобок и используется, например, при умножении двухзначных чисел.

Пример: $23 \cdot 4 = (20+3)4 = 80 + 12 = 92$.

Второе тождество иллюстрирует прием, называемый *вынесением за скобку*.

Пример: $3 \cdot 6 + 7 \cdot 6 = (3+7)6 = 10 \cdot 6 = 60$.

В заключение отметим, что сумма и произведение двух натуральных чисел всегда *существуют и определены однозначно*.

Возведение в степень. Точно таким же способом, каким из сложения мы получили умножение, теперь из умножения можно получить новое действие, называемое *возведением в степень*.

Произведение, состоящее из b сомножителей, каждый из которых равен натуральному числу a, обозначают через a^b (читается: a в степени b).

Пример: $a \cdot a = a^2$, $\quad a \cdot a \cdot a = a^3$ и так далее.

Одно число можно рассматривать как произведение, состоящее из одного сомножителя, поэтому в таком случае используется обозначение $a = a^1$. Выражение a^b называется *b-й степенью числа a*, a называется *основанием степени*, b — *показателем степени*. Возведение в степень *не коммутативно*. Это значит, что a^b и b^a, *вообще говоря*, не равны, хотя в частном случае и может иметь место равенство.

Пример: $2^4 = 2 \cdot 2 \cdot 2 \cdot 2 = 16$ и $4^2 = 4 \cdot 4 = 16$.

Однако в общем случае это неверно, что можно доказать одним примером:

$$2^3 = 2 \cdot 2 \cdot 2 = 8 \quad \text{и} \quad 3^2 = 3 \cdot 3 = 9.$$

Возведение в степень не обладает и свойством ассоциативности.

Пример: $(2^2)^3 = 4^3 = 64$ и $2^{(2^3)} = 2^8 = 256$.

В то же время имеет место тождество $(a^b)^c = a^{bc}$. В самом деле, $(a^b)^c$ обозначает произведение c сомножителей, каждое из которых есть a^b и, следовательно, в свою очередь является произведением b сомножителей, каждое из которых есть a. Таким образом a присутствует как сомножитель $b \cdot c$ раз, то есть рассматриваемое произведение действительно равно a^{bc}. Полученное тождество и на этот раз перепишем дважды:

$$\boxed{(a^b)^c = a^{bc}; \qquad a^{bc} = (a^b)^c.}$$

Первая запись выражает следующее: *при возведении степени в степень показатели перемножаются.* Согласно второй записи, *чтобы возвести число a в степень bc, достаточно возвести его в степень b, а затем полученный результат возвести в степень c.*

Пример: $2^6 = (2^3)^2 = 8^2 = 64$. Если же знать степени числа 2, то, например, $32^2 = (2^5)^2 = 2^{10} = 1024$.

Обозначая возведение степени в степень, первую степень всегда ставят в скобки, поэтому запись a^{b^c} *всегда обозначает* степень $a^{(b^c)}$. При возведении степени в степень можно получать очень большие числа. Например, самое большое число, которое может быть записано с помощью трех цифр и знаков арифметических действий, есть 9^{9^9}. Это число состоит более чем из 369 миллионов знаков. Если считать по 0,5 см на каждый знак, то запись этого числа покрыла бы расстояние от Будапешта до Лондона.

Возведение в степень обладает свойствами дистрибутивного характера. Поскольку коммутативность не выполняется, то есть основание и показатель степени нельзя поменять местами, то следует рассмотреть два типа дистрибутивности возведения в степень как относительно сложения, так и относительно умножения. С этой целью рассмотрим выражения a^{bc}, $(ab)^c$, a^{b+c} и $(a+b)^c$. (Здесь мы записали степени уже с учетом того, что *основание* — будь то сумма или произведение — *следует заключать в скобки*, а *показатель — не обязательно.)* Первое выражение уже было рассмотрено выше, с последним мы встретимся позже при рассмотрении формулы Ньютона. Во втором же случае действительно имеет место дистрибутивность. В самом деле, $(ab)^c$ обозначает произведение c сомножителей, каждое из которых есть ab. Если переставить сомножители местами, и сгруппировать отдельно все a и все b, то получится как раз произведение $a^c \cdot b^c$ (например, $(ab)^3 = ab \cdot ab \cdot ab = aaa \cdot bbb = a^3 \cdot b^3$. Таким образом

$$\boxed{(ab)^c = a^c \cdot b^c \qquad \text{или} \qquad a^c \cdot b^c = (ab)^c.}$$

Следовательно, степень произведения равна произведению степеней сомножителей с тем же показателем. Обратно: *произведение степеней с одинаковым показателем можно получить, если в ту же степень возвести произведение оснований.*

Выражение a^{b+c} обозначает произведение $b+c$ сомножителей, каждое из которых есть a. Разбив произведение на две группы по b и c сомножителей, мы получим произведение $a^b \cdot a^c$ (например, $a^{3+2} = a^5 = aaaaa = (aaa)(aa) = a^3 \cdot a^2$). Таким образом

$$\boxed{a^{b+c} = a^b \cdot a^c, \qquad a^b \cdot a^c = a^{b+c}.}$$

Итак, *если показатель степени есть сумма, то степень равна произведению степеней с тем же основанием и с показателями, равными слагаемым этой суммы*. Обратно: *при умножении степеней с одинаковыми основаниями показатели складываются.*

Здесь также следует отметить, что любое натуральное число можно возвести в степень с любым натуральным показателем, при этом результат определен однозначно.

Рассмотренные выше действия: сложение, умножение и возведение в степень — называют *прямыми действиями*. В случае, если известен результат прямого действия и требуется найти одно из «составляющих», мы имеем дело с *обратным действием*.

Вычитание. *Действие, обратное сложению, называется вычитанием.* Пусть известна сумма и одно из слагаемых, требуется найти второе слагаемое. (Поскольку сложение коммутативно, то все равно, которое из слугаемых требуется найти). Другими словами, для данных натуральных чисел a и b требуется найти такое натуральное число x, чтобы выполнялось равенство $a + x = b$. Такое натуральное число можно найти не всегда. Например, если $a = 5$, $b = 3$, то такого натурального числа найти нельзя. В случае же $a = 3$ и $b = 5$ подходит число $x = 2$, и это единственное подходящее число. Аналогично положение и в общем случае: *если a и b — произвольные натуральные числа, то или существует такое (однозначно определенное) натуральное число x, для которого $a + x = b$ или существует такое (однозначно определенное) натуральное число y, для которого $b + y = a$, или же $a = b$, то есть возможен один из трех перечисленных случаев.*

Если для натуральных чисел a и b найдется такое натуральное число x, что $a + x = b$, то это число определено числами a и b однозначно, поэтому для него можно использовать обозначение $b - a$. Число $b - a$ называется *разностью* между *уменьшаемым* b и *вычитаемым* a, само же действие называется *вычитанием*. Согласно сказанному выше, в области натуральных чисел вычитание можно выполнить не всегда, но если можно, то результат определен однозначно. Вычитание, естественно, не коммутативно. С дальнейшими свойствами вычитания мы познакомимся при рассмотрении целых чисел.

Неравенства. Если существует натуральное число $b - a$, то говорят, что b больше a или a меньше b и обозначают этот факт как $b > a$ или $a < b$. Эти знаки называются знаками неравенства. Перечеркнутый знак неравенства означает, что соответствующее неравенство не выполняется, например: $3 < 5$ или $5 > 3$, но $5 \not< 3$ и $3 \not> 5$. Как мы видели при рассмотрении вычитания, *для любых натуральных чисел a и b выполняется одно и только одно из соотношений*

$$a < b, \quad a = b, \quad a > b.$$

Это свойство называют *трихотомией* неравенства. Нетрудно убедиться в том, что отношение неравенства обладает свойством *транзитивности:* если $a > b$ и $b > c$, то $a > c$.

Ниже приведены соотношения, выражающие связь между неравенствами и действиями над числами. Если a, b и c — *натуральные числа,* и $a > b$, то

$$\boxed{\begin{array}{llll} a + c > b + c > b, & a \cdot c > b \cdot c > b, & a^c > b^c; & \\ \text{если} & c > 1, & \text{то} & c^a > c^b. \end{array}}$$

Комбинируя эти соотношения, можно получать новые. Приведем важнейшее из них:

если $a > b$ и $c > d$ — натуральные числа, то

$$a + c > b + d \quad \text{и} \quad a \cdot c > b \cdot d.$$

Деление. *Действие, обратное умножению, называется делением.* Деление означает нахождение одного из сомножителей, когда известны произведение и другой сомножитель. Если для натуральных чисел a и b найдется такое натуральное число x, что $a \cdot x = b$, то можно доказать, что это число единственное. Такое однозначно определенное число обозначают через $\frac{b}{a}$, при этом b называют *делимым*, a — *делителем*, а $\frac{b}{a}$ — *частным*. Например, $\frac{6}{2} = 3$. В области натуральных чисел деление, вообще говоря, не выполнимо. Не существует, например, натурального числа, которое при умножении на 2 дало бы 3. Если деление $\frac{b}{a}$ удается выполнить (в области натуральных чисел), то число b называют *кратным* числу a, а число a — *делителем* числа b. (Таким образом, слово «делитель» используется здесь в двух смыслах: в первом случае это то *число, на которое делят,* во втором — *число, на которое делится* некоторое другое число). С дальнейшими свойствами деления мы познакомимся при рассмотрении рациональных чисел.

Извлечение корня и логарифмирование. При рассмотрении действия, обратного возведению в степень, следует учесть, что последнее не коммутативно. По этой причине для возведения в степень существует два обратных действия.

Если известны степень и ее показатель, и требуется найти основание, то говорят об извлечении корня. Если же известны степень и ее основа-

ние и требуется найти показатель, то мы имеем дело с логарифмированием.

В области натуральных чисел эти действия, вообще говоря, не выполнимы. Например, никакая натуральная степень числа 2 не равна трем; ни одно натуральное число во второй степени не равно пяти. К этим двум действиям мы вернемся при рассмотрении вещественных чисел.

ЦЕЛЫЕ ЧИСЛА

Как мы видели, в области натуральных чисел обратные действия, вообще говоря, не выполнимы. Тем не менее, во многих задачах это необходимо.

Рассмотрим, например, следующую задачу: сыну a лет, а его отцу b лет. *Через* сколько лет отец будет вдвое старше сына? Предположим, что это произойдет через x лет. Тогда сыну будет $a + x$ лет, а отцу $b + x$ лет. По условию в это время возраст отца будет вдвое больше возраста сына, то есть $b + x = 2(a + x)$. Используя свойства действий с помощью простых преобразований, получаем: $x = b - 2a$. Например, при $b = 30$ и $a = 10$ получается, что $x = 30 - 2 \cdot 10 = 10$. (Через 10 лет сыну будет 20 лет, а отцу 40, то есть действительно вдвое больше, чем сыну). Однако если $b = 40$, $a = 20$, то $x = 40 - 40$; если же $b = 50$, $a = 30$, то $x = 50 - 60$. Такое вычитание уже *нельзя выполнить в области натуральных чисел*. В то же время на вопрос задачи *можно* ответить, так как в первом случае отец как раз *сейчас* вдвое старше сына, а во втором случае был вдвое старше сына 10 лет *назад*.

При решении таких задач возникает необходимость в числах, которые выразили бы понятия «сейчас» или «тому назад». Другими словами *необходимо ввести новые числа; множество чисел необходимо расширить*. При этом мы хотим получить такое множество чисел, которое было бы шире множества натуральных чисел, и в котором можно было бы решать задачи «типа» 40—40 и 50—60. (Мы не сомневаемся, что все читатели этой книги знакомы с числом 0 и с отрицательными числами, но хотим показать, *почему и как* нужно было ввести эти числа.) Мы хотим определить новые числа таким образом, чтобы

1. *среди них были все натуральные числа;*
2. *в области новых чисел всегда можно было бы выполнить вычитание;*
3. *действия, определенные в области новых чисел, если их применять к натуральным числам, давали бы те же результаты, что и соответствующие действия, определенные ранее в области натуральных чисел.*

Целесообразно ввести новые числа таким образом, чтобы остались в силе свойства действий, с которыми мы познакомились при рассмотрении натуральных чисел. Здесь, однако, возникает проблема. Дело в том, что в области натуральных чисел можно было выполнить лишь одно из вычитаний: $b - a$ или $a - b$. В области новых чисел из-за условия 2 это будет не так. По этой причине речь может идти только о том, чтобы «как можно больше свойств осталось в силе», что называют *принципом постоянства* действий. Итак, последним условием является

4. *принцип постоянства* (действий).

Поскольку в области натуральных чисел нельзя произвести вычитания $1-1$, то результат этого действия будет новым числом, которое мы обозначим через 0 и будем называть *нулем*. Согласно перечисленным условиям, для каждого натурального числа a должно существовать такое число $(-a)$, для которого $a+(-a)=0$. Если мы хотим, следуя условию 4, чтобы разность была определена однозначно, то числа $0, (-1), (-2), \ldots$ должны быть различными (доказательства этого факта мы не приводим). Числа $1, 2, 3, \ldots, 0, (-1), (-2), (-3), \ldots$ называются *целыми числами*. Рассматривая натуральные числа как часть целых чисел, мы называем их *положительными целыми числами*, в то время как числа $(-1), (-2), (-3), \ldots$ называются *отрицательными целыми числами*.

ДЕЙСТВИЯ НАД ЦЕЛЫМИ ЧИСЛАМИ

Сложение. Поскольку для каждого слагаемого и для суммы существуют три возможности (они могут быть положительными, равными 0 или отрицательными), то следует рассмотреть отдельно 9 случаев. (Если мы *хотим* определить сложение таким образом, чтобы выполнялась коммутативность, то достаточно 6 случаев). Нет необходимости отдельно рассматривать случай, когда оба слагаемых — положительные числа, так как тогда, согласно условию 3, результат *должен* совпадать с результатом сложения, определенного в области натуральных чисел.

Рассмотрим случай, когда одно из слагаемых есть 0, а другое положительно. Итак, чему должна быть равна сумма $a+0$ (a — положительное целое)? Если мы хотим, чтобы в области целых чисел выполнялось свойство ассоциативности сложения, то *должно выполняться равенство* $(a+0)+1=a+(0+1)$. Число 0 мы определили равенством $0=1-1$, но это означает, что $0+1=1$. Следовательно, $(a+0)+1=a+1$. *Если мы хотим*, чтобы разность и здесь была определена однозначно, то отсюда должно следовать, что $a+0=a$.

Итак, *если мы хотим, чтобы указанные выше свойства выполнялись и в области целых чисел, то сложение необходимо определить таким образом, чтобы выполнялось тождество* $a+0=0+a=a$. *Отсюда, естественно,* еще не следует, что указанные свойства выполняются *всегда*.

Аналогичным образом нетрудно убедиться, что *если принять во внимание принцип постоянства, то в области целых чисел сложение может быть определено только следующим образом:*

$$0+(-a)=(-a)+0=(-a);\quad 0+0=0;\quad (-a)+(-b)=-(a+b);$$

$$a+(-b)=(-b)+a=\begin{cases} a-b, & \text{если } a>b,\\ 0, & \text{если } a=b,\\ -(b-a), & \text{если } a<b.\end{cases}$$

Так, например, $0+(-3)=(-3)$; $(-3)+(-5)=(-8)$; $5+(-3)=2$, $5+(-5)=0$, $3+(-5)=(-2)$.

Можно показать, что если определить сложение таким образом, то для любых двух целых чисел сумма существует и определена однозначно; кроме того, имеет место коммутативность и ассоциативность сложения (доказательство этих фактов потребовало бы трудоемких и кропотливых вычислений). Из определения видно, что вместо $a + (-b)$ можно писать $a - b$, а выражение $(-a) + (-b)$ может быть записано как $-a - b$, поэтому и вместо $(-a)$ можно писать $-a$.

Для любого *целого* числа a вместо $0 - a$ можно писать просто $-a$. Тогда очевидно, что $-(-a) = a$, например, $-(-3) = 3$. Иногда вместо a пишут $+a$. *Это не означает, что a положительно.* Так, например, $+(-3) = -3$ не является положительным числом. И, наконец, можно показать, что для любых двух целых чисел существует и однозначно определена их разность (в заданном порядке). Заметим также, что любое целое число может быть представлено в виде разности двух положительных целых чисел, например, $6 = 7-1$, $0 = 1-1$ и $-5 = 1-6$.

Умножение. Посмотрим теперь, как следует *целесообразно* определить умножение целых чисел. Если мы хотим, чтобы умножение было дистрибутивным относительно сложения, то, умножив обе части равенства $0 + 0 = 0$, на произвольное число a, мы должны получить: $a(0 + 0) = a \cdot 0 + a \cdot 0 = a \cdot 0$. Если теперь из обеих частей последнего равенства вычесть $a \cdot 0$, то получится, что $a \cdot 0 = 0$. Вообще можно показать, что *если имеет место дистрибутивность*, то должны выполняться следующие тождества:

$$a0 = 0a = 0, \quad a(-b) = (-a)b = -ab, \quad (-a)(-b) = ab,$$

где a, b — целые числа.

Пример: $3 \cdot 0 = 0, \quad (-3)5 = -15, \quad (-7)(-4) = 28.$

Можно показать, что при таком определении умножения действительно имеют место коммутативность, ассоциативность и дистрибутивность.

Возведение в степень. Для целых чисел можно говорить и о возведении в степень. Пока *выражение a^b имеет для нас смысл только тогда, когда b — натуральное число*. Поскольку a^b обозначает произведение b сомножителей, каждый из которых равен a, то, сохраняя определение умножения, нетрудно видеть, что *любая степень числа* 0 *есть* 0; *четная степень отрицательного числа положительна, нечетная — отрицательна*. Если в показатель степени записать 0 или отрицательное число, то прежнее определение степени теряет смысл, так как нет смысла говорить о произведении, содержащем 0 или отрицательное число сомножителей.

Вычитание. Вернемся опять к свойствам вычитания. С помощью простых вычислений нетрудно доказать следующие тождества:

$$a + (b - c) = (a + b) - c, \quad a - (b - c) = (a - b) + c,$$

$$a - (b + c) = (a - b) - c.$$

Пример: $5 + (-3) = 5 + (1 - 4) = (5 + 1) - 4 = 6 - 4 = 2,$

$$5 - 3 = 5 - (4 - 1) = (5 - 4) + 1 = 1 + 1 = 2,$$

$$7 - 5 = 7 - (2 + 3) = (7 - 2) - 3 = 5 - 3 = 2.$$

Легко доказывается и тождество

$$\boxed{(a - b)c = ac - bc.}$$

Например:

$$(-3)2 = (2 - 5)2 = 2 \cdot 2 - 5 \cdot 2 = 4 - 10 = -6.$$

В области целых чисел можно говорить и об *упорядочении* (то есть о том, что одно число больше или меньше другого). Говорят, что целое число a больше целого числа b ($a > b$), если разность $a - b$ положительна. Так, например, $-3 < -2 < -1 < 0 < 1 < 2 < 3$. Здесь также имеют место трихотомия и транзитивность. Однако связь неравенств с арифметическими действиями в области целых чисел уже иная. Если, например, обе части неравенства $5 > 3$ умножить на -1, то получаемое формально неравенство $-5 > -3$ уже неверно, а наоборот, $-5 < -3$. Вообще можно показать, что *неравенство не нарушится, если к обеим его частям прибавить одно и то же число или если обе его части умножить на одно и то же положительное число*. Однако *если обе части неравенства умножить на одно и то же отрицательное число, то неравенство превращается в неравенство противоположного смысла* (то есть большей станет та часть неравенства, в которой до умножения стояло меньшее число, как в приведенном примере). Соответствующим образом изменится и связь неравенств с возведением в степень, однако рассматривать ее, как правило, нет необходимости.

Деление. В области целых чисел деление, вообще говоря, не выполнимо. Если же его удается выполнить, то частное не обязательно определено однозначно. Например, 0 «можно разделить» на 0, то есть можно найти такое целое число x, что $0 \cdot x = 0$, только беда в том, что существует не одно такое число — любое целое число удовлетворяет этому равенству. Таким образом, результат деления оказывается не однозначным, в то же время *мы хотим, чтобы результаты всех действий были определены однозначно*. Действиями, обратными возведению в степень, мы займемся лишь при рассмотрении вещественных чисел.

РАЦИОНАЛЬНЫЕ ЧИСЛА

Аналогично тому, как знание натуральных чисел оказалось недостаточным для решения некоторых задач, так для ряда задач оказывается мало и целых чисел. Чтобы можно было выполнять деление, потребовалось ввести новые, так называемые *рациональные числа*. Их название происходит от латинского слова *ratio*, что означает *отношение*. Другими словами, дело не в том, что эти числа имеют смысл, а иррациональные

— в противоположность им — не имеют смысла. Причиной такого названия послужило то, что древние греки с помощью этих чисел обозначали отношение длин соизмеримых отрезков и хотя не считали их числами, все же производили с ними вычисления как с числами. Если имеется два отрезка, длины которых равны 2 и 3 единицам измерения, то отношение этих длин обозначают как 2 : 3 или $\frac{2}{3}$. В то же время это отношение не зависит от выбора единицы измерения. Если, например, за единицу измерения принять вдвое меньший отрезок, то отношение тех же длин будет выражено как $\frac{4}{6}$. Однако, поскольку оба «числа» обозначают одно и то же отношение, то должно выполняться равенство $\frac{2}{3} = \frac{4}{6}$. Таким образом, *эти отношения могут быть равны и в случае различных данных.* Принимая во внимание, что никакой отрезок нельзя соизмерить с отрезком длины 0, приходим к выводу, что «число, стоящее внизу», не может быть равно 0. Итак, желание определить отношение (частное) любых двух целых чисел приводит нас к таким числам-отношениям.

Рациональными числами называются выражения вида $\frac{a}{b}$, где a и b — целые числа, причем $b \neq 0$. Рациональные числа $\frac{a}{b}$ и $\frac{c}{d}$ равны тогда и только тогда, когда $ad = bc$ (тем самым равенство между рациональными числами сводится к равенству между целыми числами, которые можно считать известными).

Число a называется *числителем*, а b — *знаменателем* рационального числа $\frac{a}{b}$, разделяющая их черта называется *чертой дроби*. Рациональные числа называют также *дробями*. (Название числителя и знаменателя появилось из следующих соображений: знаменатель *знаменует* собой, то есть определяет, *какую часть* единицы измерения следует рассмотреть, числитель же указывает, *сколько* таких частей нужно взять).

ОБЩИЕ СВЕДЕНИЯ О РАЦИОНАЛЬНЫХ ЧИСЛАХ

Сложение и умножение этих новых чисел можно определять по-разному. Однако можно показать, что если мы хотим, чтобы числа вида $\frac{a}{1}$ «вели себя» точно так же, как целые числа, и чтобы имели место коммутативность и ассоциативность действий, а также дистрибутивность умножения относительно сложения, то сложение и умножение можно определить только следующим образом:

$$\boxed{\frac{a}{b} + \frac{c}{d} = \frac{ad + bc}{bd}; \qquad \frac{a}{b} \cdot \frac{c}{d} = \frac{ac}{bd}.}$$

(Заметим, что данное определение было бы единственно возможным и в том случае, если бы мы не потребовали выполнения всех перечисленных условий, а только некоторых из них).

Можно доказать, что сложение и умножение, определенные таким образом, обладают свойствами коммутативности и ассоциативности, и кроме того, выполняется дистрибутивность умножения относительно сложения.

Рассматривая сумму и произведение чисел, знаменатели которых равны 1, получаем: $\frac{a}{1}+\frac{c}{1}=\frac{a+c}{1}$, $\frac{a}{1}\cdot\frac{c}{1}=\frac{ac}{1}$, что действительно соответствует сложению и умножению, определенным в области целых чисел.

Правило возведения рациональных чисел в степень сразу получается из определения умножения, если учесть, что возведение в степень есть «повторенное умножение». Показатель степени и здесь может быть только натуральным числом.

Не представляет затруднений и вычитание рациональных чисел. Поскольку вычитание есть действие, обратное сложению, то получаем, что

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

Нетрудно убедиться, что разность между любыми двумя рациональными числами всегда существует и определена однозначно.

При делении, однако, уже возникают проблемы, хотя именно невыполнимость деления привела к необходимости ввести дроби. Мы уже видели, что при делении «0 на 0» результат получается не однозначным. Кроме того, выясняется, что *никакое число* (кроме 0) *нельзя делить на* 0. Действительно, разделить число $\frac{a}{b}$ на 0 — это значит найти такое число $\frac{x}{y}$, которое при умножении на 0 давало бы $\frac{a}{b}$. Но если $\frac{a}{b}\neq 0$, то такого числа не существует, так как любое число при умножении на 0 даст 0. Во всех остальных случаях в области рациональных чисел деление выполняется однозначно. Другими словами, *при делении произвольного рационального числа на любое другое рациональное число, отличное от нуля, получается однозначный результат.*

Многие считают чуть ли не недостатком математики тот факт, что на 0 делить нельзя. В то же время речь идет лишь о том, что определенные свойства взаимно исключают друг друга, предположение же об их одновременной выполнимости приводит к противоречию. В отдельных случаях такие «взаимоисключения» очевидны, в других же — как например, при делении на 0 — к такому выводу можно прийти лишь путем некоторых рассуждений. Если бы кто-нибудь все-таки проявил упорство и попытался бы ввести такие «числа», в области которых действия производились бы по обычным правилам, и, кроме того, было бы осуществимо и деление на 0, то получилось бы по пословице: «За двумя зайцами погонишься, ни одного не поймаешь». В самом деле, можно показать, что множество чисел, удовлетворяющее всем этим требованиям, состоит из одного нуля.

В области рациональных чисел можно определить и упорядочение. *Рациональное число* $\frac{a}{b}$ *положительно*, если a и b — положительные целые числа. Говорят, что рациональное число $\frac{a}{b}$ больше рационального числа $\frac{c}{d}$, $\left(\frac{a}{b} > \frac{c}{d}\right)$, если разность $\frac{a}{b} - \frac{c}{d}$ есть положительное рациональное число. Здесь справедливы все те соотношения, с которыми мы познакомились при рассмотрении целых чисел.

В обиходе рациональные числа чаще называют дробями или *обыкновенными дробями* (в отличие от десятичных дробей). Ниже мы рассмотрим действия и «правила» действий над дробями.

Всякая дробь может быть записана в виде, где в знаменателе стоит натуральное число. По определению равенства между рациональными числами $\frac{a}{b} = \frac{-a}{-b}$. Поскольку одно из чисел b или $-b$ положительно (так как $b \neq 0$), то в знаменателе одной из двух записанных дробей стоит положительное целое число. Если знаменатель положителен, то дробь уже имеет искомый вид. Если же знаменатель отрицательный, то знаки перед числителем и знаменателем нужно поменять на противоположные.

Пример: $\frac{15}{-7} = \frac{-15}{7}$; $\frac{-9}{-7} = \frac{9}{7}$.

Сокращение и «расширение» дробей. Из определения равенства между рациональными числами следует, что *величина дроби не изменится, если и числитель, и знаменатель умножить или — если деление выполнимо — разделить на одно и то же целое число, отличное от* 0. (При этом числитель и знаменатель дроби, естественно, изменятся).

Пример: $\frac{5}{7} = \frac{5 \cdot 3}{7 \cdot 3} = \frac{15}{21}$; $\frac{12}{16} = \frac{3 \cdot 4}{4 \cdot 4} = \frac{3}{4}$.

Приведение дробей к общему знаменателю. Умножая числитель и знаменатель на одно и то же число, можно любые две дроби привести к виду, где они имеют *одинаковые знаменатели*, например: $\frac{5}{7}$ и $\frac{3}{5}$. Умножая числитель и знаменатель первой дроби на 5, а второй — на 7, получаем: $\frac{5}{7} = \frac{5 \cdot 5}{7 \cdot 5} = \frac{25}{35}$ и $\frac{3}{5} = \frac{3 \cdot 7}{5 \cdot 7} = \frac{21}{35}$.

Вообще говоря, всегда приводит к цели *умножение числителя и знаменателя каждой из дробей на знаменатель другой дроби.* Однако, чтобы избежать появления больших чисел, целесообразно выбирать возможно меньший общий знаменатель. *Наименьшим из возможных общих знаменателей является наименьшее общее кратное двух знаменателей* (см. «Теория чисел», стр. 176).

Пример: Наименьшее общее кратное знаменателей дробей $\frac{17}{42}$ и $\frac{23}{35}$ есть $[42, 35] = 210$. Поскольку $210 = 5 \cdot 42 = 6 \cdot 35$, то дроби можно привести к общему знаменателю следующим образом:

$$\frac{17}{42} = \frac{17 \cdot 5}{42 \cdot 5} = \frac{85}{210} \text{ и } \frac{23}{35} = \frac{23 \cdot 6}{35 \cdot 6} = \frac{138}{210}.$$

ДЕЙСТВИЯ НАД РАЦИОНАЛЬНЫМИ ЧИСЛАМИ (ДРОБЯМИ)

Сложение и вычитание. *Сложение (вычитание) дробей с одинаковыми знаменателями.* Из определения сложения и равенства между дробями следует, что *числитель суммы (разности) дробей с одинаковыми знаменателями равен сумме (разности) числителей этих дробей, а знаменатель — их общему знаменателю.*

Пример: $\frac{3}{7} + \frac{5}{7} = \frac{8}{7}$, $\frac{15}{31} - \frac{-7}{31} = \frac{15-(-7)}{31} = \frac{15+7}{31} = \frac{22}{31}$.

Сложение (вычитание) дробей с разными знаменателями. На практике сложение и вычитание дробей производят следующим образом: *сначала дроби приводят к общему знаменателю, а затем производят сложение (вычитание) дробей с одинаковыми знаменателями.*

Пример: Используя результаты предыдущего примера, получаем:

$$\frac{5}{7} - \frac{3}{5} = \frac{25}{35} - \frac{21}{35} = \frac{4}{35}; \quad \frac{17}{42} + \frac{23}{35} = \frac{85}{210} + \frac{138}{210} = \frac{223}{210}.$$

Умножение. *Умножение дроби на целое число.* Из определения умножения и сокращения дробей следует, что *произведение дроби на целое число можно получить, если числитель дроби умножить на это число, а знаменатель оставить прежним. Если же это целое число является делителем знаменателя дроби, то мы получим тот же результат, если знаменатель разделим на это число, а числитель оставим прежним.*

Пример: $\frac{3}{7} \cdot 2 = \frac{3 \cdot 2}{7} = \frac{6}{7}$; $\frac{5}{6} \cdot 2 = \frac{5}{6:2} = \frac{5}{3}$. В последнем случае мы по сути дела произвели сокращение дроби: $\frac{5}{6} \cdot 2 = \frac{5 \cdot 2}{6} = \frac{5 \cdot 2}{3 \cdot 2} = \frac{5}{3}$.

Выполнение умножения вторым способом позволяет избежать появления больших чисел. Можно было бы дать и общий метод для избежания больших чисел, но формулировка оказалась бы слишком сложной, поэтому мы выскажем лишь основной принцип. Если умножать дробь на целое число первым способом, то перед тем, как перемножить числа, стоящие в знаменателе, следует посмотреть, не является ли полученная дробь сократимой.

Пример: $\frac{89}{21}\cdot 14=\frac{89\cdot 14}{21}=\frac{89\cdot 2\cdot 7}{3\cdot 7}=\frac{89\cdot 2}{3}=\frac{178}{3}$.

Умножение дроби на дробь. Из определения умножения непосредственно следует, что *произведение двух дробей есть дробь, числитель которой равен произведению числителей, а знаменатель — произведению знаменателей сомножителей.*

Пример: $\frac{35}{78}\cdot\frac{72}{91}=\frac{35\cdot 72}{78\cdot 91}=\frac{2520}{7098}$.

Здесь также можно постараться избежать появления больших чисел, если перед умножением чисел, стоящих в числителе и знаменателе, посмотреть, не является ли полученная дробь сократимой.

В приведенном выше примере имеем: $\frac{35\cdot 72}{78\cdot 91}=\frac{7\cdot 5\cdot 6\cdot 12}{6\cdot 13\cdot 7\cdot 13}=\frac{5\cdot 12}{13\cdot 13}=$ $=\frac{60}{169}$.

Дробь, обратная данной. При делении дробей существенную помощь оказывает понятие дроби, обратной данной.

Дробью, обратной дроби $\frac{a}{b}$ (если $a\neq 0$) называется дробь $\frac{b}{a}$. Так, например, для дроби $\frac{2}{3}$ обратной является дробь $\frac{3}{2}$; для дроби $\frac{-5}{4}$ — дробь $\frac{4}{-5}=\frac{-4}{5}$.

Число, обратное целому числу, можно получить, записав это число в виде дроби со знаменателем 1. Например, записав число 2 в виде $\frac{2}{1}$ видим, что обратным числу 2 является $\frac{1}{2}$. Вообще числом, обратным целому числу a, является $\frac{1}{a}$.

Деление. *Деление дроби на дробь.* С помощью понятия обратной дроби деление дробей можно свести к умножению: *частное от деления дроби на дробь равно произведению делимого на число, обратное делителю.*

Пример: $\frac{2}{3}:\frac{5}{7}=\frac{2}{3}\cdot\frac{7}{5}=\frac{14}{15}$.

Деление дроби на целое число. Как частный случай из правила деления дроби на дробь следует, что *частное от деления дроби на целое число можно получить, умножив на делитель знаменатель или разделив на него (если возможно деление нацело) числитель дроби.*

Пример: $\frac{3}{7}:5=\frac{3}{7\cdot 5}=\frac{3}{35}$, $\frac{6}{7}:3=\frac{6:3}{7}=\frac{2}{7}$.

Возведение в степень. *Возведение дроби в степень.* Пользуясь определением степени, возведение дроби в степень с положительным целым показателем можно свести к умножению дробей. *Чтобы получить степень дроби с положительным целым показателем, достаточно разделить соответствующую степень числителя на степень знаменателя.*

Пример: $\left(\frac{2}{3}\right)^3=\frac{2^3}{3^3}=\frac{8}{27}$.

(При рассмотрении перечисленных правил в интересах краткости и доступности мы иногда пренебрегали строгостью изложения.)

В связи с дробями упомянем еще три устаревшие понятия, которые сравнительно редко, но употребляются математиками.

«*Правильная дробь*». Так называют положительные дроби, меньшие единицы. Название указывает на то, что число действительно является частью единицы.

«*Неправильная дробь.*» Так называются все положительные дроби, не являющиеся правильными.

«*Смешанное число.*» Так называют сумму положительного целого числа и правильной дроби. Это понятие позволяет лучше оценить величину неправильной дроби.

Пример: $\frac{48}{7}=\frac{42+6}{7}=\frac{42}{7}+\frac{6}{7}=6+\frac{6}{7}$.

Смешанное число иногда записывают в виде как $6\frac{6}{7}$, однако, такая запись неудобна, поскольку ее легко перепутать с произведением чисел 6 и $\frac{6}{7}$.

ПЕРИОДИЧЕСКИЕ ДЕСЯТИЧНЫЕ ДРОБИ

Системы мер, используемые на практике, основаны, как правило, на десятичной системе счисления. Это обстоятельство придает особую важность дробям, знаменатели которых представляют собой степени числа 10. (Если, например, некоторые данные измеряют в метрах, то целесообразно рассматривать дециметр как десятую, а сантиметр — как сотую часть метра.)

Дроби, знаменатели которых являются некоторыми степенями числа 10, называются конечными десятичными дробями.

Десятичные дроби чаще обозначают следующим образом: записывают только числитель дроби и, *если знаменатель есть k-ая степень 10-ти, то перед k-ым десятичным знаком числителя (считая от единиц, то есть справа) ставят запятую.*

Пример: $\frac{315}{10}=31{,}5$ (читается: 31 целая 5 десятых); $\frac{17}{100}=0{,}17$ (читается: 0 целых, 17 сотых); $\frac{-5721}{1000}=-5{,}721$ (читается: минус 5 целых, 721 тысячная). Если числитель делится на 10, то дробь сначала сокращают: $\frac{270}{100} = \frac{27}{10} = 2{,}7$; это соответствует отбрасыванию нуля в конце десятичной дроби: $\frac{270}{100} = 2{,}70 = 2{,}7$. Этот процесс записи соответствует делению числителя дроби на знаменатель (дробь $\frac{a}{b}$ действительно является результатом деления a на b): В самом деле

$$\frac{3417}{1000} = \frac{3000+400+10+7}{1000} = \frac{3000}{1000} + \frac{400}{1000} + \frac{10}{1000} + \frac{7}{1000} =$$

$= 3 + \frac{4}{10} + \frac{1}{100} + \frac{7}{1000} = 3{,}417$ (то есть 3 целых, 4 десятых, 1 сотая, 7 тысячных).

В виде конечной десятичной дроби могут быть записаны те и только те дроби, знаменатели которых не содержат других простых множителей, кроме 2 и 5 (или лишь один из них).

В самом деле, если в знаменателе дроби стоит некоторая степень 10-ти, то после сокращения в знаменателе остается число, которое не может иметь других простых множителей, кроме тех, которые содержатся в 10-ти, то есть кроме 2 и 5.

Пример: $3{,}25 = \frac{325}{100} = \frac{13 \cdot 25}{4 \cdot 25} = \frac{13}{4}$; $5{,}12 = \frac{512}{100} = \frac{128 \cdot 4}{25 \cdot 4} = \frac{128}{25}$.

Обратно: если в разложении знаменателя на простые множители присутствуют лишь числа 2 и 5, то знаменатель является произведением некоторой степени 2-х на некоторую степень 5-ти. Тогда после умножения числителя и знаменателя дроби на соответствующую степень 2-х или 5-ти в знаменателе окажется некоторая степень 10-ти.

Пример: $\frac{317}{125} = \frac{317 \cdot 8}{125 \cdot 8} = \frac{2536}{1000} = 2{,}536$; $\frac{3}{8} = \frac{3 \cdot 125}{8 \cdot 125} = \frac{375}{1000} = 0{,}375$.

Таким образом, никакие другие дроби не могут быть представлены в виде конечной десятичной дроби. Однако известные преобразования можно совершать и с другими дробями.

Чему, например, равняется $\frac{10}{7}$ метра? Выражая это расстояние в метрах, получаем, что $\frac{7+3}{7} = 1 + \frac{3}{7}$ метра, то есть это расстояние меньше 2 м и больше 1 м. Для большей точности посмотрим, сколько это приблизительно дециметров: $\frac{10}{7}$ м $= \frac{100}{7}$ дм. $\frac{100}{7} = \frac{98+2}{7} = 14 + \frac{2}{7}$ дм. Значит, данное расстояние больше 14 дм и меньше 15 дм. Однако выше мы уже выяснили, что это расстояние больше 1 м и меньше 2 м, так что достаточно было рас-

смотреть остаток, равный $\frac{3}{7}$ м. Учитывая это, продолжим процесс дальше:
$$\frac{10}{7}\text{ м} = 1\text{ м} + \frac{3}{7}\text{ м} = 1\text{ м} + \frac{30}{7}\text{ дм} = 1\text{ м} + 4\text{ дм} + \frac{2}{7}\text{ дм} = 1{,}4\text{ м} + \frac{20}{7}\text{ см} =$$
$$= 1{,}4\text{ м} + 2\text{ см} + \frac{6}{7}\text{ см} = 1{,}42\text{ м} + \frac{60}{7}\text{ мм} = 1{,}42\text{ м} + 8\text{ мм} + \frac{4}{7}\text{ мм} =$$
$$= 1{,}428\text{ м} + \frac{4}{7}\text{ мм} = \ldots$$

Этот процесс в точности соответствует процессу деления, с той лишь разницей, что на каждом шагу мы переходили к единице измерения, в десять раз меньшей, то есть всякий раз умножали делимое на 10 или, что то же самое, приписывали к нему справа 0. Таким образом, «деление» производилось по следующей схеме:

```
10 : 7 = 1,428571 . . .
 30
  20
   60
    40
     50
      10
       3
        .
         .
          .
```

Отсюда видно, что в процессе деления в остатке снова появляется 3. С этого места весь процесс повторится. Из этого сразу следует, что процесс бесконечен. С другой стороны, следующие десятичные знаки частного можно записать, не производя деления: 1,42857142 . . .
Рассмотрим еще два примера:

```
31 : 11 = 2,818181 . . .     и     53 : 12 = 4,416666 . . . .
 90                                  50
  20                                  20
   90                                  80
    2                                   8
     .                                   .
      .                                   .
       .                                   .
```

Нетрудно заметить, что, начиная с некоторого места, цифры появляются группами. Это происходит потому, что рано или поздно одно из чисел появляется в остатке вторично. Действительно, при делении на некоторое натуральное число b после $b + 1$ шагов мы получим $b + 1$ остатков. Все эти остатки не могут быть разными, так как при делении на b в остатке могут получаться лишь числа $0, 1, \ldots, b - 1$, то есть число возможных остатков равно b (а из b чисел нельзя выбрать $b + 1$ *различных*).

Итак, путем деления *из всякой дроби можно получить бесконечную последовательность цифр, в которой, начиная с некоторого места, повторяется некоторая группа цифр. Такие «числа» называют периодическими десятичными дробями. Если повторяющаяся группа цифр появляется сразу после запятой, то получается чистая, в противном же случае — смешанная периодическая десятичная дробь.*

Целое число, стоящее перед запятой, называется *целой частью*, а число, состоящее из цифр повторяющейся группы, называется *периодом* дроби.

Пример: В рассмотренных выше случаях имеем:

Периодическая десятичная дробь	целая часть	число, стоящее до первого периода	период
1,428571 ...	1	отсутствует	428571
2,81 ...	2	отсутствует	81
4,416 ...	4	41	6

Период дроби достаточно записать один раз. Однако, чтобы не перепутать получаемые таким образом дроби с конечными десятичными дробями, над первой и последней цифрами периода ставят точку (если период состоит из одной цифры, то достаточно поставить одну точку). Такое обозначение одновременно позволяет выделить и число, стоящее до первого периода*. (В некоторых книгах для обозначения периода над ним проводят черту.) Таким образом для рассмотренных выше периодических дробей имеем:

$$1{,}\dot{4}2857\dot{1}\,; \qquad 2{,}\dot{8}\dot{1}\,; \qquad 4{,}41\dot{6}.$$

Можно доказать, что *не существует обыкновенной дроби, преобразовав которую в десятичную дробь, мы получили бы период, состоящий из одних девяток.* По этой причине такие числа и *не считаются десятичными дробями.*

Поставим теперь обратный вопрос: для всякой ли периодической десятичной дроби найдется обыкновенная дробь, которая в нее преобразуется? Ответ и на этот раз оказывается положительным. (Для доказательства не обязательно использовать бесконечную геометрическую прогрессию; искомую обыкновенную дробь можно задать, исходя из свойств деления. Доказательство этого достаточно длинно, поэтому мы не станем его излагать, а приведем лишь формулу, задающую дробь). Очевидно, что целую часть дроби можно при этом не принимать во внимание, в конце ее можно просто прибавить к полученному результату. Таким образом, достаточно рассмотреть случай дробей, меньших единицы. *Пусть P обозначает число, стоящее до первого периода, Q — период, а PQ — число, стоящее до второго периода дроби. Тогда обыкновенную дробь, которая преобразуется в данную*

* Другое обозначение: заключение в круглые скобки периода дроби. Так, рассмотренные выше числа будут иметь вид: 1, (428571); 2, (81) и 4,41 (6). (Прим.ред.

периодическую десятичную дробь, можно получить следующим образом:

$$\boxed{\dfrac{PQ - P}{\underbrace{9\ldots9}\ \underbrace{0\ldots0}}}$$

$\uparrow$ столько девяток, сколько цифр в периоде;

$\uparrow$ столько нулей, сколько цифр в числе, стоящем до первого периода.

В рассмотренных выше примерах: $\dfrac{428571}{999999} = \dfrac{11 \cdot 13 \cdot 37 \cdot 81}{7 \cdot 11 \cdot 13 \cdot 27 \cdot 37} = \dfrac{3}{7}$ $\left(\text{действительно } 1 + \dfrac{3}{7} = \dfrac{10}{7}\right)$; $\dfrac{81}{99} = \dfrac{9}{11}$ $\left(\text{и } 2 + \dfrac{9}{11} = \dfrac{31}{11}\right)$; $\dfrac{416 - 41}{900} = \dfrac{375}{900} = \dfrac{5}{12}$ $\left(\text{и } 4 + \dfrac{5}{12} = \dfrac{53}{12}\right)$.

К периодическим десятичным дробям можно отнести и все конечные десятичные дроби как дроби, не имеющие периода. Чтобы выяснить, не производя преобразования, будет ли периодическая десятичная дробь смешанной или чистой, достаточно рассмотреть знаменатель обыкновенной дроби. Для этого сначала нужно сделать дробь *несократимой*, то есть чтобы числитель и знаменатель были взаимно простыми числами. Из основной теоремы теории чисел следует, что любая обыкновенная дробь приводится к несократимому виду однозначно. Чтобы привести дробь к несократимому виду, нужно произвести все возможные сокращения. Можно доказать, что путем деления *всякую обыкновенную дробь можно преобразовать в периодическую десятичную дробь* (период которой не может состоять из одних девяток), и *из всякой периодической десятичной дроби* (с помощью указанного выше процесса) *можно получить* (однозначно определенную) *обыкновенную дробь, которая в нее преобравуется. Между несократимыми дробями и соответствующими периодическими дробями существует следующая связь:*

Если в периодической десятичной дроби:		тогда в разложении знаменателя соответствующей несократимой дроби на простые множители,
отсутствует период	$\leftrightarrow$	*могут присутствовать только 2 и 5,*
первый период начинается сразу после запятой	$\leftrightarrow$	*все множители отличны от 2 и 5,*
первый период начинается не сразу после запятой	$\leftrightarrow$	*присутствует хотя бы одно из чисел 2 и 5, а также другие числа.*

ВЕЩЕСТВЕННЫЕ ЧИСЛА

Рациональные числа можно изображать в виде точек на прямой линии (так называемой *числовой оси*). Выберем на горизонтальной прямой точку, которую примем за начало отсчета, эта точка будет изображать число 0. Зададим теперь отрезок единичной длины. Точка прямой, находящаяся справа от точки 0 на расстоянии, равном длине единичного отрезка, будет изображать число 1. На расстоянии в 2, 3, ... единицы длины справа от точки 0 отметим точки, изображающие, соответственно, числа 2, 3, ... На таком же расстоянии слева от точки 0 изобразим, соответственно, числа -1, -2, -3, ... Аналогичным образом изображаются на числовой оси и дроби. Например, изображение дроби $\frac{3}{7}$ можно получить, трижды отложив справа от точки 0 седьмую часть единичного отрезка. Итак, каждому рациональному числу соответствует некоторая точка числовой оси. Однако обратное утверждение неверно : на числовой оси есть точки, которым не соответствует никакое рациональное число. Такую точку можно получить, если, например, отложить из точки 0 диагональ квадрата с единичной стороной. Таким точкам можно поставить в соответствие так называемые иррациональные числа. Первым из иррациональных чисел стало известно число π (половина длины окружности единичного радиуса). Однако тот факт, что это число — иррациональное, не могли доказать до 1882 года. По-видимому впервые такой факт был установлен для числа $\sqrt{2}$ (равного длине диагонали единичного квадрата). Все рациональные и иррациональные числа вместе называются вещественными (или действительными) числами.

Вещественные числа мы введем точно таким же способом, каким были введены целые и рациональные числа. С помощью рациональных чисел мы построим новые математические объекты, обладающие всеми теми свойствами, которыми мы хотим наделить вещественные числа. Такой метод мы называем методом создания модели.

Однако здесь возникают новые и очень существенные трудности. До сих пор каждое новое число нам удавалось определить с помощью *конечного числа* ранее известных чисел. Для задания вещественных чисел требуется, вообще говоря, бесконечно много рациональных чисел. Именно поэтому было дано много различных способов введения вещественных чисел (см. гл. «Математический анализ», стр. 381). Мы изложим здесь основную идею метода, предложенного Дедекиндом и использующего так называемые *сечения* в области рациональных чисел.

Между любыми двумя иррациональными числами на числовой оси найдется рациональное число. Отсюда следует, что тем самым это иррациональное число будет определено однозначно, если задать, какие из рациональных чисел больше него и какие меньше. Другими словами, всякое иррациональное число однозначно разбивает множество рациональных чисел на два подмножества так, что

1. *каждое рациональное число оказывается элементом одного и только одного из двух подмножеств;*

2. *одно из подмножеств таково, что всякое относящееся к нему рациональное число больше любого рационального числа из другого подмножества (первое из подмножеств называется верхним, а второе — нижним классом сечения).*
(Для однозначности сечения предполагают также, что в верхнем классе нет наименьшего числа.)

Можно показать, что для любого сечения, удовлетворяющего указанным свойствам, всегда существует такое вещественное число (рациональное или иррациональное), которое меньше любого рационального числа из верхнего класса, но не меньше любого рационального числа из нижнего класса. Рассмотрим, например, следующее сечение в области рациональных чисел: пусть к верхнему классу принадлежат все положительные рациональные числа, квадраты которых больше 2-х, а к нижнему — отрицательные рациональные числа и те, квадраты которых меньше 2. Нетрудно видеть, что таким образом мы действительно получаем сечение, удовлетворяющее указанным выше свойствам, и это сечение определяет число $\sqrt{2}$.

Вещественные числа могут быть заданы и в виде так называемых *бесконечных десятичных дробей*. Эти дроби отличаются от периодических десятичных дробей тем, что не обязательно имеют период. Точнее: бесконечная десятичная дробь тогда и только тогда задает некоторое рациональное число, когда является периодической. *Среди вещественных чисел не может быть такого числа, в записи которого, начиная с некоторого места, повторялась бы лишь цифра* 9 (действительно, такое число было бы периодическим, в то же время такой периодической десятичной дроби не существует).

ДЕЙСТВИЯ НАД ВЕЩЕСТВЕННЫМИ ЧИСЛАМИ

В области вещественных чисел также можно определить действия, причем таким образом, что, будучи примененными к рациональным числам, они совпадают с действиями, известными до сих пор.

Сложение и умножение. *В области вещественных чисел сложение и умножение всегда выполнимы и их результаты определены однозначно. Оба действия коммутативны и ассоциативны; кроме того, имеет место дистрибутивность умножения относительно сложения.* В случае натурального показателя возведение в степень может быть выполнено, как «повторяющееся» умножение.

Вычитание. Вычитание также всегда выполнимо однозначно, и обладает всеми свойствами, с которыми мы познакомились при рассмотрении целых чисел.

Вещественное число называется положительным, если нижний класс соответствующего ему сечения содержит положительное рациональное число (то есть положительно всякое число, «большее» какого-либо положительного рационального числа).

Можно показать, что

а) *если ни вещественное число α, ни $-\alpha$ не являются положительными, то* $\alpha = 0$. Это означает, что если вещественное число α отлично от 0, то либо α

положительно, либо положительно число $-\alpha$. В последнем случае число α называют *отрицательным*.

б) *Сумма и произведение положительных вещественных чисел также положительны.*

В области вещественных чисел можно ввести упорядочение, как это было сделано для целых чисел, при этом остаются в силе все рассмотренные там соотношения для неравенств.

Нетрудно показать, что *между любыми двумя вещественными числами найдется рациональное число*, точнее: каковы бы ни были вещественные числа $\alpha < \beta$, найдется такое рациональное число r, что $\alpha < r < \beta$.

Абсолютная величина. Рассмотрим изображение вещественных чисел на числовой оси. Расстояние от точки 0 до произвольной точки числовой оси мы назовем абсолютной величиной соответствующего вещественного числа. Абсолютная величина вещественного числа α обозначается символом $|\alpha|$. Очевидно, что только абсолютная величина нуля равна 0. Нетрудно видеть, что $|\alpha| = \alpha$, если α положительно, и $|\alpha| = -\alpha$, если α отрицательно. Таким образом, абсолютная величина произвольного вещественного числа, отличного от 0, положительна. Абсолютная величина числа удовлетворяет следующим двум соотношениям:

$$\boxed{|\alpha\cdot\beta| = |\alpha|\cdot|\beta|; \qquad |\alpha+\beta| \leqq |\alpha|+|\beta|.}$$

(Знак $\leqq$ (или $\geqq$) означает, что число, стоящее слева (справа) не больше числа, стоящего справа (слева), то есть «меньше или равно»).

Деление. В области вещественных чисел деление также всегда выполнимо однозначно (за исключением *деления на* 0).

Возведение в степень. Начнем с установления связи между возведением в степень и упорядочением в области вещественных чисел. Из правил умножения сразу следует, что *четная степень отрицательного числа положительна, нечетная степень — отрицательна*. Очевидно также, что $(-\alpha)^2 = \alpha^2$. Далее, *если $\alpha > \beta$ — положительные вещественные числа, а n — произвольное натуральное число, то $\alpha^n > \beta^n$*. Это утверждение может быть доказано с помощью полной индукции. Рассмотрим лишь случай $n \leqq 3$, отсюда станет ясен весь процесс доказательства. При $n = 1$ доказываемое утверждение справедливо, так как совпадает с условием. В случае $n = 2$ произведем следующие преобразования: $\alpha^2 - \beta^2 = \alpha^2 - \alpha\beta + \alpha\beta - \beta^2 = \alpha(\alpha-\beta) + \beta(\alpha-\beta)$. Поскольку сомножители в обоих слагаемых положительны, то оба слагаемые, а следовательно и вся сумма также положительны, что и доказывает требуемое неравенство. Для $n = 3$ произведем аналогичные преобразования: $\alpha^3-\beta^3 = \alpha^3-\alpha\beta^2 + \alpha\beta^2-\beta^3 = \alpha(\alpha^2-\beta^2) + \beta^2(\alpha-\beta)$. Учитывая, что второй сомножитель первого слагаемого положителен по только что доказанному, получаем, что и в этом случае все сомножители, а с ними оба произведения и вся сумма также положительны.

Извлечение корня. Перейдем к рассмотрению *извлечения корня*, одного из действий, обратных возведению в степень. О понятии извлечения корня мы упоминали, когда требовалось по данной степени и данному показателю определить основание степени. Поскольку при возведении в степень показатель всегда был натуральным числом, то же самое должно выполняться и при извлечении корня. Соотношение $\alpha^n = \beta$ мы будем обозначать и как $\alpha = \sqrt[n]{\beta}$ (читают: корень n-ой степени из бэта). Итак, для любого натурального числа n и любого вещественного числа a, $\sqrt[n]{a}$ обозначает такое вещественное число, n-я степень которого равна a. Число a называют при этом *подкоренным числом*, n — *показателем* корня, $\sqrt[n]{a}$ — *корнем* или *радикалом*. В случае $n = 1$ очевидно, что $\sqrt[1]{a} = a$, поэтому для этого случая нет необходимости вводить отдельное обозначение. Для $n = 2$ показатель корня не записывают:

$$\sqrt{a} = \sqrt[2]{a}$$

(читают: корень квадратный из a).

При извлечении корня уже возникает немало проблем. Дело в том, что извлечь корень удается не всегда, а если и удается, то не всегда однозначно. Рассмотрим сначала случай, когда *под корнем* стоит *положительное число*. *Если a — положительное число, то существует одно и только одно положительное число b такое, что $b^n = a$.* (В этом нетрудно убедиться с помощью сечений в области рациональных чисел. Действительно, разобьем множество рациональных чисел на два подмножества, к одному из которых отнесем все рациональные числа, n-я степень которых меньше a, а к другому — все остальные. В результате — если рассмотреть только положительные числа — мы получим некоторое сечение в области рациональных чисел, которое определяет ровно одно положительное вещественное число). Заметим, что такой метод позволяет определить значение корня с любой степенью точности.

Пример: Рассмотрим $\sqrt{2}$. Поскольку $1^2 = 1 < 2$, а $2^2 = 4 > 2$, то $1 < \sqrt{2} < 2$. Далее, $1{,}4^2 = 1{,}96 < 2$, но $1{,}5^2 = 2{,}25 > 2$, следовательно, $1{,}4 < \sqrt{2} < 1{,}5$. Аналогичным образом $1{,}41^2 = 1{,}9881 < 2$, но $1{,}42^2 = 2{,}0164 > 2$, следовательно $1{,}41 < \sqrt{2} < 1{,}42$ и так далее.

Мы видели, что квадрат (2-ая степень) любого вещественного числа совпадает с квадратом того же числа, взятого с обратным знаком. Отсюда получаем, что если n — четное, то из $b^n = a$ следует $(-b)^n = a$. (В то же время очевидно, что других таких чисел не существует. Таким образом, в этом случае извлечение корня оказывается «двузначным действием». Чтобы избежать этого, условимся считать, что *в случае положительного a корень $\sqrt[n]{a}$ всегда положителен. Так, например,* $\sqrt{9} = 3$, второе же число, квадрат которого равен 9, будем обозначать как $-\sqrt{9}$ $(=-3)$.

Поскольку четная степень отрицательного числа положительна, то *из отрицательного числа нельзя извлечь корня четной степени;* можно, однако, показать, что *корень нечетной степени из отрицательного числа есть число отрицательное* (например, $\sqrt[3]{-27} = -3$).

Принцип монотонности. Используя тот факт, что для положительных чисел из $a > b$ следует, что $a^n > b^n$, можно доказать, что *для положительных чисел a и b из $a > b$ следует, что $\sqrt[n]{a} > \sqrt[n]{b}$*. Выражая эти два факта, говорят, что *возведение в степень и извлечение корня обладают свойством монотонности*.

Свойства извлечения корня. Эти свойства получаются из соответствующих свойств возведения в степень и по существу совпадают с ними.

Для степени:	Соответствующее тождество для корня:
$(a^n)^k = a^{nk}$	$\sqrt[n]{(\sqrt[k]{a})} = \sqrt[nk]{a}$.
$(ab)^n = a^n \cdot b^n$	$\sqrt[n]{a \cdot b} = \sqrt[n]{a} \cdot \sqrt[n]{b}$.
$\left(\dfrac{a}{b}\right)^n = \dfrac{a^n}{b^n}$	$\sqrt[n]{\dfrac{a}{b}} = \dfrac{\sqrt[n]{a}}{\sqrt[n]{b}}$.

Свойства корней также имеют двойной смысл, согласно тому, «с какой стороны рассматривать соответствующее тождество». Первое из записанных тождеств можно выразить словами следующим образом: *чтобы извлечь корень из корня, достаточно перемножить их показатели и извлечь из подкоренного числа корень полученной степени*. Обратно: *если в показателе корня стоит произведение, то значение корня можно получить, если из подкоренного числа извлечь сначала корень степени, равной одному из сомножителей, а из полученного результата извлечь корень степени, равной второму сомножителю*.

Пример: $\sqrt[6]{64} = \sqrt[6]{8^2} = \sqrt[3]{\sqrt[2]{8^2}} = \sqrt[3]{8} = 2$ ($\sqrt{8^2} = 8$ по определению.)

Следующие два расшифровываются аналогично.

Корень из произведения равен произведению корней той же степени из сомножителей. Обратно: *произведение корней с равными показателями равно корню той же степени из произведения подкоренных чисел*.

Пример: $\sqrt[3]{10} \cdot \sqrt[3]{100} = \sqrt[3]{1000} = 10$.

Корень из дроби равен частному от деления корня той же степени из числителя на корень той же степени из знаменателя дроби. Обратно (вместо подробной и точной формулировки скажем коротко): *«отношение корней равно корню из отношения»*.

Пример: $\dfrac{\sqrt[3]{120}}{\sqrt[3]{15}} = \sqrt[3]{\dfrac{120}{15}} = \sqrt[3]{8} = 2.$

В частном случае второе из записанных выше тождеств принимает следующий вид:

$$\boxed{\sqrt[n]{a^n \cdot b} = a \cdot \sqrt[n]{b} \quad \text{или} \quad a \cdot \sqrt[n]{b} = \sqrt[n]{a^n \cdot b}.}$$

В первом случае говорят о *вынесении множителя из-под корня*, во втором — *о внесении под корень*.

Пример: $\sqrt{8} = \sqrt{2^3} = \sqrt{2^2 \cdot 2} = 2 \cdot \sqrt{2}$ или $5 \cdot \sqrt{2} = \sqrt{25 \cdot 2} = \sqrt{50}.$

ОБОБЩЕНИЕ ПОНЯТИЯ СТЕПЕНИ

Со степенями какого-нибудь фиксированного числа очень просто производить вычисления. Рассмотрим, например, степени числа 2:

показатель:	1	2	3	4	5	6	7	8	9	10	11	12	13
степень:	2	4	8	16	32	64	128	256	512	1024	2048	4096	8192

Отсюда получаем, например, что $16 \cdot 128 = 2^4 \cdot 2^7 = 2^{11} = 2048$ или $16^3 = (2^4)^3 = 2^{12} = 4096$. Таким образом, вместо умножения можно производить сложение, а вместо возведения в степень — умножение. Для этого, правда, нужна таблица, аналогичная приведенной выше. Однако эта таблица имеет недостаток, состоящий в том, что в нижней строке присутствуют далеко не все числа. Так, например, умножение $3 \cdot 7$ уже не удается произвести указанным выше способом (в данном случае это, конечно, не вызывает затруднений, так как эти числа можно просто перемножить в уме. Однако очень часто положение бывает иным. Еще более существенную помощь указанный способ может оказать при вычислении сложных выражений, когда необходимо произвести несколько действий). Таким образом, появилась необходимость выразить все числа в виде степеней некоторого фиксированного числа. Однако с помощью одних лишь натуральных показателей это неосуществимо. Если, например, число a меньше единицы, то число $\dfrac{a+a^2}{2}$ наверняка лежит между a и a^2, а следовательно, не может быть целой степенью числа a. Вследствие этого *понятие степени нужно было обобщить* таким образом, чтобы появилась возможность рассматривать степени не только с целым показателем. Но это обобщение нельзя было произвести на основе прежнего определения степени, так как произведение может состоять лишь из натурального числа сомножителей. Именно поэтому *нельзя знать заранее, останутся ли свойства степени справедливыми и в общем случае.* В то же время это необходимо, так как *упрощение вычислений становится возможным как раз благодаря этим свойствам.*

Итак, с одной стороны, то, *как следует трактовать степень в общем случае, зависит от определения.* С другой стороны, *нет смысла обобщать понятие степени таким образом, чтобы не оставались справедливыми прежние свойства.* Но тогда вовсе не обязательно, что желаемое обобщение осуществимо. (В математике тоже нередко случается, когда что-то «хорошо было бы» сделать — но не получается. Вспомним, например, деление на 0).

Обобщение понятия степени происходит в несколько этапов. Сначала определим степень с показателем 0 и с целым отрицательным показателем — тем самым будет определено понятие степени с произвольным *целым показателем.* После этого мы определим степень с дробным, а значит, и с произвольным *рациональным показателем.* В заключение будет определена степень с любым *вещественным показателем.* Напомним, что условие сохранения прежних свойств по существу эквивалентно принципу постоянства действий.

Пусть для любого вещественного числа a

$$a^0 = 1, \qquad a^{-n} = \frac{1}{a^n},$$

где n — натуральное число.

С помощью несложных выкладок нетрудно убедиться, что при данном определении для любого вещественного $a \neq 0$ справедливы следующие соотношения:

$$\boxed{a^{nk} = (a^n)^k; \qquad a^{n+k} = a^n \cdot a^k,}$$

где n и k — любые целые числа. Кроме того, если b — произвольное вещественное число, отличное от 0, то выполняются также соотношения

$$\boxed{(ab)^n = a^n \cdot b^n; \qquad \left(\frac{a}{b}\right)^n = \frac{a^n}{b^n}.}$$

Данное выше определение не «взято с потолка», больше того, из принципа постоянства следует, что это единственное целесообразное определение. В самом деле, если мы хотим, чтобы соотношение $a^{n+k} = a^n \cdot a^k$ выполнялось для любых целых n и k, то при $k = 0$ и при $k = -n$ получаем:

$$a^n = a^{n+0} = a^n \cdot a^0; \qquad a^0 = a^{n+(-n)} = a^n \cdot a^{-n}.$$

Первое тождество — в случае $a \neq 0$ — выполняется только если $a^0 = 1$, поэтому целесообразным является именно это определение. Если это определение уже принято, то из второго тождества можно установить целесообразность определения степени с целым отрицательным показателем. Отсюда, конечно, еще не следует, что требуемые свойства выполняются всегда; это требует дополнительного доказательства.

Рассмотрим несколько примеров:

$$2^{-2} = \frac{1}{2^2} = \frac{1}{4}; \qquad 3^{-1} = \frac{1}{3};$$

$$\left(\frac{1}{2}\right)^{-3} = \frac{1}{\left(\frac{1}{2}\right)^3} = \frac{1}{\frac{1}{8}} = 8 \qquad \left(\text{или} \quad \left(\frac{1}{2}\right)^{-3} = (2^{-1})^{-3} = 2^3 = 8\right).$$

Для $a = 0$ отрицательная степень не имела бы смысла, так как на 0 делить нельзя. В случае, когда показатель степени равен 0, также возникает проблема: свойства степени выполняются и при $0^0 = 0$, и при $0^0 = 1$. Ввиду этого если где-то приходится использовать степень 0^0, то ее определяют одним из двух способов, в зависимости от того, который из них удобней.

Пусть для любого положительного вещественного числа a

$$a^{\frac{p}{q}} = \sqrt[q]{a^p},$$

где p — целое, а q — натуральное число.

Всякое рациональное число может быть записано в виде $\frac{p}{q}$ так, чтобы в знаменателе стояло положительное число, поэтому данное определение распространяется на случай степени с произвольным рациональным показателем. При этом следует учесть, что равенство $\frac{p}{q} = \frac{k}{n}$ возможно и в случае, если $p \neq k$ и $q \neq n$. Можно показать, что в этом случае $a^{\frac{p}{q}} = a^{\frac{k}{n}}$. Нетрудно убедиться и в справедливости соотношений

$$\boxed{a^{rs} = (a^r)^s, \qquad a^{r+s} = a^r \cdot a^s; \qquad (ab)^r = a^r \cdot b^r; \qquad \left(\frac{a}{b}\right)^r = \frac{a^r}{b^r},}$$

где a и b — положительные вещественные числа, а r и s — рациональные числа. Целесообразность данного определения также вытекает из принципа постоянства. Если мы хотим, чтобы тождество $a^{rs} = (a^r)^s$ выполнялось для всех рациональных r и s, то оно должно выполняться и для $r = \frac{p}{q}$, $s = q$, где p — целое, а q — натуральное число. Но тогда $a^p = a^{\frac{p}{q} \cdot q} = \left(a^{\frac{p}{q}}\right)^q$, поэтому если $b = a^{\frac{p}{q}}$, то должно выполняться тождество $b^q = a^p$, откуда по определению корня следует, что $b = \sqrt[q]{a^p}$. Это и означает целесообразность данного определения. Естественно, что выполнение соответствующих свойств степени и здесь нуждается в доказательстве.

Рассмотрим несколько примеров степеней с дробным показателем:

$$8^{\frac{2}{3}} = \sqrt[3]{8^2} = \sqrt[3]{64} = 4; \qquad 8^{\frac{2}{3}} = (2^3)^{\frac{2}{3}} = 2^{3\cdot\frac{2}{3}} = 2^2 = 4;$$

$$9^{-\frac{3}{2}} = 9^{\frac{-3}{2}} = \sqrt{9^{-3}} = \sqrt{\frac{1}{9^3}} = \sqrt{\frac{1}{729}} = \frac{1}{\sqrt{729}} = \frac{1}{27} \quad \text{или}$$

$$9^{-\frac{3}{2}} = (3^2)^{-\frac{3}{2}} = 3^{2\cdot\left(-\frac{3}{2}\right)} = 3^{-3} = \frac{1}{3^3} = \frac{1}{27};$$

$$\left(\frac{1}{32}\right)^{-\frac{3}{5}} = \left(\frac{1}{32}\right)^{\frac{-3}{5}} = \sqrt[5]{\left(\frac{1}{32}\right)^{-3}} = \sqrt[5]{32^3} = \sqrt[5]{32768} = 8 \quad \text{или}$$

$$\left(\frac{1}{32}\right)^{-\frac{3}{5}} = (2^{-5})^{-\frac{3}{5}} = 2^{(-5)\cdot\left(\frac{-3}{5}\right)} = 2^3 = 8.$$

Степень с основанием 0 мы исключили из рассмотрения уже в случае целого показателя. В случае же дробного показателя приходится исключить и степень с отрицательным основанием. В противном случае мы имели бы, например, что $(-1)^{1/2} = -1$, в то же время не существует вещественного числа, квадрат которого был бы числом отрицательным. Можно было бы, правда, попытаться допустить отрицательное основание в случае, когда в знаменателе показателя стоит нечетное число. Однако если бы это и удалось, то вряд ли оказалось бы удобным, так как всякий раз (в том числе и при исследовании свойств степени) приходилось бы отдельно смотреть, допустимо ли в данном случае отрицательное основание. Выясняется однако, что степень с отрицательным основанием не допустима и в этом случае. Действительно, так как, например, $(-1)^{1/3} = \sqrt[3]{-1} = -1$ и в то же время $(-1)^{1/3} = (-1)^{2/6} = \sqrt[6]{(-1)^2} = \sqrt[6]{1} = 1$, то из однозначности степени следовало бы равенство $-1 = 1$, что невозможно.

Посмотрим, можем ли мы записать любое положительное вещественное число в виде рациональной степени некоторого *фиксированного* положительного числа. Если да, то можно показать, что в качестве такого фиксированного числа можно выбрать число 2. Однако это не так.

Например, число 3 так записать не удается. Действительно, если бы выполнялось равенство $2^{\frac{p}{q}} = 3$, то при положительном p должно выполняться равенство $2^p = 3^q$, что невозможно, поскольку слева стоит четное, а справа — нечетное число. (Слева может стоять нечетное число, но лишь при $p = 0$, но тогда и q должно равняться 0, в то время как в знаменателе не может стоять 0). Таким образом, необходимо ввести понятие степени и с не рациональными показателями.

Для обобщения понятия степени на случай произвольного вещественного показателя мы воспользуемся следующими двумя свойствами степеней с рациональным показателем:

1. *если вещественное число $a > 1$ и r — положительное рациональное число, то $a^r > 1$;*

2. *если вещественное число $a > 1$, и $r > s$ — рациональные числа, то $a^r > a^s$.*

2-е свойство можно сформулировать следующим образом: *при возрастании показателя рациональная степень числа, большего единицы, монотонно возрастает.*

Случай, когда основание меньше единицы, можно свести к предыдущему. Действительно, если $a < 1$, то $a^{-1} > 1$, и положительные степени числа a совпадают с отрицательными степенями числа a^{-1}. Отсюда следует, что в этом случае при возрастании показателя степень уменьшается. В случае $a = 1$ для любого рационального числа r имеем: $a^r = 1$.

Теперь, принимая во внимание монотонность, можно дать определение степени вещественного числа, большего единицы, с произвольным вещественным показателем (а значит, и определение степени, основание которой положительно, но меньше единицы). При этом мы будем опираться на следующий факт:

пусть a и x — вещественные числа, причем $a > 1$. Тогда существует такое и притом единственное вещественное число b, что для любых рациональных r и s

$$\text{из } r < x < s \qquad \text{следует, что } a^r < b < a^s.$$

Очевидно, что теперь *по определению* следует положить $b = a^x$. Можно доказать, что при таком определении выполняются все свойства степени; в случае же, когда x — рациональное число, так определенная степень совпадает со степенью, получаемой согласно прежнему определению. Кроме того, очевидно, что *степень обладает свойством монотонности*, то есть (при $a > 1$) *из $x > y$ следует, что $a^x > a^y$.*

В качестве примера найдем, чему равна степень $2^{\sqrt{3}}$. О точном определении, естественно, не может идти речи. Известный нам метод позволяет лишь определить два числа, между которыми заключено значение степени, хотя эти границы можно определить с любой степенью точности. Поскольку $1 = 1^2 < 3 < 4 = 2^2$, то $1 < \sqrt{2} < 2$, откуда $2^1 < 2^{\sqrt{3}} < 2^2$, то есть $2 < 2^{\sqrt{3}} < 4$. Так как $1{,}7^2 = 2{,}89 < 3 < 3{,}24 = 1{,}8^2$, то $1{,}7 < \sqrt{3} < 1{,}8$, и значит $2^{1,7} < 2^{\sqrt{3}} < 2^{1,8}$. Как видим, здесь уже представляет трудность и определение самих чисел, задающих границы. Самое большее, что можно сделать, это следующее: поскольку $1{,}5 < \sqrt{3}$, то $2^{1,5} = \sqrt{8} < 2^{\sqrt{3}}$; так как $2{,}8^2 = 7{,}84 < 8$, то $2{,}8 < 2^{\sqrt{3}}$.

ПОНЯТИЕ ЛОГАРИФМА

Как мы видели, возведение в степень не является коммутативным действием. По этой причине существует два (различных) действия, обратных действию возведения в степень. Одно из них — извлечение корня, с которым мы подробно познакомились выше. Обобщение понятия степени позволило и извлечение корня рассматривать как возведение в степень (показатель которой есть число, обратное показателю корня).

Другое действие, обратное возведению в степень, — это отыскание логарифма или логарифмирование. При рассмотрении степени мы видели, что если a — положительное, а x — произвольное вещественное число, то $b = a^x$ — однозначно определенное положительное число. Логарифмирование заключается в отыскании показателя степени, когда известны сама степень и ее основание.

Соотношение $b = a^x$ в так называемом «логарифическом виде» записывают следующим образом: $x = \log_a b$ (читают: логарифм b по основанию a). Иногда используют также обозначения $x = {}^a\log b$ или $x = \underset{a}{\underbrace{\log}}\, b$. Число a называется *основанием логарифма* (или просто *основанием*), само же число x называют *логарифмом* числа b по основанию a. Итак, ***при возведении числа a в степень*** $\log_a b$ ***мы получаем*** b.

Как мы уже видели, логарифм имеет смысл лишь для положительного числа при положительном основании. Если $a = 1$, то и $b = 1$, поэтому предполагается также, что основание не равно единице.

По определению имеем, например:

$$\log_2 8 = 3, \text{ так как } 2^3 = 8,$$

$$\log_2 \frac{1}{2} = -1, \text{ так как } 2^{-1} = \frac{1}{2},$$

$$\log_8 \frac{1}{4} = -\frac{2}{3}, \text{ так как } 8^{-\frac{2}{3}} = \frac{1}{4}.$$

Свойства логарифма. С помощью логарифмирования удается значительно упростить многие (приближенные) вычисления. Это оказывается возможным благодаря следующим тождествам:

$$1.\ \log_a (bc) = \log_a b + \log_a c; \qquad 2.\ \log_a \frac{b}{c} = \log_a b - \log_a c;$$

$$3.\ \log_a b^d = d \cdot \log_a b,$$

где a, b и c — положительны, а d — произвольное вещественное число. Эти тождества получаются из соответствующих тождеств для степени. В качестве примера докажем первое из них, два другие доказываются аналогично. Пусть $x = \log_a(bc)$, $u = \log_a b$ и $v = \log_a c$. По определению логарифма имеем: $bc = a^x$, $b = a^u$ и $c = a^v$. Тогда из свойств степени получаем, что $a^x = bc = a^u \cdot a^v = a^{u+v}$. По свойству монотонности степени отсюда следует, что $x = u + v$, что и требовалось доказать.

Сформулируем указанные выше свойства и на словах. Итак, *при данном (фиксированном) основании* имеем:

1. *Логарифм произведения равен сумме логарифмов сомножителей.* Обратно: *сумма логарифмов двух чисел равна логарифму их произведения.*

2. *Логарифм дроби равен разности между логарифмами числителя и знаменателя.* Обратно: *разность логарифмов равна логарифму дроби, числитель которой равен первому логарифмируемому числу, а знаменатель — второму.*

3. *Логарифм степени равен произведению показателя степени на логарифм основания.* Обратно: *произведение какого-либо числа на логарифм некоторого числа равно логарифму степени с основанием, равным второму числу, и показателем, равным первому числу.*

Пример:

$$\log_6 4 + \log_6 9 = \log_6 36 = 2;$$

$$\log_4 5 - \log_4 20 = \log_4 \frac{1}{4} = -1;$$

$$6 \cdot \log_{16} 4 = \log_{16} 4^6 = \log_{16} 4096 = 3.$$

Корень логарифмируется как степень с дробным показателем.

Системы логарифмов. Принимая во внимание, что все перечисленные тождества имеют место для некоторого фиксированного основания, обычно рассматривают логарифмы с фиксированным основанием. Чаще всего используются два основания. В математике очень важную роль играют так называемые *натуральные логарифмы*, основание которых обозначают буквой e. Это число является одной из важнейших постоянных (так же, как и число π). Число e является иррациональным, $e = 2{,}71828\ldots$ Логарифмы по основанию e обозначают через ln (иногда используется также обозначение «log»), символ «ln» читают как «логарифм натуральный». Для практических целей чаще используют логарифмы по основанию 10, так как 10 лежит в основе употребительной системы счисления. *Логарифм по основанию 10* обозначается символом lg (в некоторых книгах используется также обозначение «log»).

НОРМАЛЬНАЯ ФОРМА ЧИСЕЛ, ХАРАКТЕРИСТИКА И МАНТИССА

В этом параграфе речь будет идти о том, почему используют логарифмы по основанию 10. Рассмотрим целые степени числа 10:

$$\ldots,\ \frac{1}{1000},\ \frac{1}{100},\ \frac{1}{10},\ 1,\ 10,\ 100,\ 1000,\ \ldots$$

Нетрудно видеть (и доказать), что всякое положительное число заключено между двумя целыми степенями числа 10 (или равно одной из них). Например, для числа 0,037 имеем: $0{,}01 < 0{,}037 < 0{,}1$; для 0,1 выполняются неравенства $0{,}1 \leq 0{,}1 < 1$; для числа 1967 получаем: $1000 < 1967 < 10000$. Вообще для любого положительного вещественного числа a найдется такое (однозначно определенное) целое число n, что $10^n \leq a < 10^{n+1}$.

Отсюда после умножения на 10^{-n} получаем: $1 \leqq a \cdot 10^{-n} < 10$. Таким образом, для числа $b = a \cdot 10^{-n}$ имеем:

$$a = b \cdot 10^n, \qquad 1 \leqq b < 10.$$

Итак, *всякое положительное число может быть представлено в виде произведения некоторого числа, заключенного между числами 1 и 10, на некоторую целую степень 10-ти. Такое представление однозначно.*

Пример: $0{,}037 = 3{,}7 \cdot 10^{-2}$; $0{,}1 = 1 \cdot 10^{-1}$; $1967 = 1{,}967 \cdot 10^3$.

Такое представление мы называем *нормальной формой числа*. Показатель степени числа 10, присутствующий в нормальной записи числа, называют *порядком* этого числа. Такая форма записи оказывается очень удобной в случае очень больших или очень малых чисел.

Если $a = b \cdot 10^n$, то $\log a = \log 10^n + \log b = n + \log b$.

Из монотонности степени следует, что и логарифм обладает свойством монотонности, то есть из неравенств $1 \leqq b < 10$ следует, что $\log 1 \leqq \leqq \log b < \log 10$. Но $\log 1 = 0$, а $\log 10 = 1$, поэтому $\log b$ всегда есть число неотрицательное, меньшее единицы. Число n называется *характеристикой*, а $\log b$ — *мантиссой* числа $\log a$. Характеристика зависит только от порядка числа, а мантисса определяется «видом» самого числа. Например, числа 0,01967, 0,1967, 1,967, 19,67, 196,7, 1967, 19670, и 196700 имеют одну и ту же мантиссу, в то время как их характеристики равны, соответственно, -2, -1, 0, 1, 2, 3, 4 и 5. Таким образом, мантиссу можно определить по виду числа, а характеристику — по положению запятой. (Причиной того, что при умножении на 10 вид числа не меняется, является использование десятичной системы счисления. Именно поэтому в качестве основания логарифмов целесообразно выбирать число 10). Нетрудно заметить (см. приведенные выше примеры), что для чисел, больших единицы, характеристика на 1 меньше количества знаков до запятой; для чисел, меньших единицы, характеристика равна числу нулей, стоящих между запятой и первой значащей цифрой, взятому со знаком минус. Согласно сказанному выше, достаточно знать логарифмы чисел от 1 до 10. Существуют таблицы, содержащие, естественно, лишь приближенные значения этих логарифмов. Однако, как правило, и данные не бывают достаточно точными. Рассмотрим пример на использование таблиц логарифмов. Пусть надо вычислить

$$x = \frac{0{,}013 \cdot 53{,}2^3}{6{,}8 \cdot \sqrt{0{,}571}}.$$

Имеем: $\log x = \log 0{,}013 + 3 \cdot \log 53{,}2 - \log 6{,}8 - \frac{1}{2} \cdot \log 0{,}571$.

$\log 0{,}013 = 0{,}1139 - 2$; $\log 53{,}2 = 1{,}7259$; $\log 6{,}8 = 0{,}8325$; $\log 0{,}571 = 0{,}7566 - 1 = 1{,}7566 - 2$ (такая запись используется для того, чтобы после деления на 2 не появилась «дробная» характеристика). Вычислим произведения: $3 \cdot 1{,}7259 = 5{,}1777$ и $\frac{1}{2} \cdot (1{,}7566-2) = 0{,}8783-1$. Таким образом, $\log x = (0{,}1139 - 2 + 5{,}1777) - (0{,}8325 + 0{,}8783 - 1) = 3{,}2916 - 0{,}7108 = 2{,}5808$, откуда $x = 389$ (конечно, приближенно).

КОМПЛЕКСНЫЕ ЧИСЛА

АЛГЕБРАИЧЕСКИЙ СПОСОБ ВВЕДЕНИЯ

В области вещественных чисел можно решить не всякое квадратное уравнение. Например, уравнение $x^2 + 1 = 0$ не имеет вещественного корня, так как не существует вещественного числа, квадрат которого равнялся бы -1. Можно, однако, «наделить» это уравнение решением. Обозначим его решение буквой i (позже все равно будет использовано это обозначение). Если это «число» присоединить к совокупности вещественных чисел, то мы получим новую систему чисел, так как «сумма и произведение чисел являются числами». Среди полученных чисел будут и числа вида $a + bi + ci^2 + \ldots + ui^n$. Однако высоких степеней числа i можно избежать, так как i должно быть числом, квадрат которого равен -1, но тогда его куб равен $-i$, четвертая степень равна 1 и так далее. Поэтому достаточно рассмотреть лишь числа вида $a + bi$. Следуя принципу постоянства действий, сложение и умножение этих «чисел» *желательно было бы* определить следующим образом:

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

$$(a + bi)(c + di) = ac + adi + bic + bdi^2 = (ac - bd) + (ad + bc)i.$$

Итак, эти «новые числа» могут быть заданы с помощью пар вещественных чисел, а действия над ними целесообразно определить указанным выше способом. Мы видели, что всякое рациональное число может быть представлено в виде различных дробей. Для «новых чисел» это неверно. В самом деле, если $a + bi = c + di$, то $(a - c) = (d - b)i$. Отсюда при $d \neq b$ получаем, что $i = \frac{a-c}{d-b}$, но это невозможно, так как i не может быть вещественным числом. Следовательно, должно быть $d = b$, откуда сразу получается, что $a = c$.

Структура «новых чисел». Итак, эти новые — так называемые *комплексные* — числа целесообразно задавать парами вещественных чисел (например, «$a + bi$» задается числами a и b) следующим образом:

$(a; b)$ — пара вещественных чисел.
$(a; b) = (c; d)$ тогда и только тогда, когда $a = c$ и $b = d$.
$(a; b) + (c; d) = (a + c; b + d)$; $(a; b)(c; d) = (ac - bd; ad + bc)$.

Можно показать, что определенные таким образом сложение и умножение коммутативны и ассоциативны и, кроме того, выполняется дистрибутивность умножения относительно сложения. Однозначно выполнимо и вычитание: $(a;\ b) - (c;\ d) = (a - c;\ b - d)$. Деление также выполняется однозначно, но это будет доказано позже.

Связь между вещественными и комплексными числами. Осталось показать, что полученная область комплексных чисел действительно шире области вещественных чисел. Другими словами, нужно убедиться, что к множеству комплексных чисел принадлежат и все вещественные числа. Но это невозможно, так как комплексные числа, в отличие от вещественных, представляют собой числовые пары. Вспомним, однако, построение множества вещественных чисел: среди них тоже не было рациональных чисел. Нам лишь удалось показать, что среди них есть числа, действия над которыми можно производить точно так же, как и над рациональными числами. То же самое мы проделаем и в данном случае. Очевидно, что вещественным числам соответствуют комплексные числа вида $(a;\ 0)$. Действительно, если каждому вещественному числу a поставить в соответствие числовую пару $(a;\ 0)$, то мы получим взаимно однозначное соответствие, которое связано с действиями следующим образом:

$$(a;\ 0) + (b;\ 0) = (a + b;\ 0);$$

$$(a;\ 0)(b;\ 0) = (ab - 0;\ 0 + 0) = (ab;\ 0).$$

Следовательно, действия с такими числовыми парами действительно можно производить точно так же, как с соответствующими вещественными числами. В первоначальном представлении $a + bi$ означало сумму, а bi — произведение, a и b обозначали вещественные числа. Рассмотрим соответствующие им комплексные числа: числу a соответствует пара $(a;\ 0)$, числу b — пара $(b;\ 0)$, а числу $i = 0 + 1i$ — пара $(0;\ 1)$. Имеем:

$$(a;\ 0) + (b;\ 0)(0;\ 1) = (a;\ 0) + (b0 - 01;\ b1 + 00) =$$
$$= (a;\ 0) + (0;\ b) = (a;\ b).$$

Таким образом, вместо комплексного числа $(a;\ 0)$ можно писать вещественное число a, так как действия с такими комплексными числами выполняются точно так же, как с соответствующими вещественными числами. Если комплексное число $(0;\ 1)$ обозначить через i, то, согласно последнему равенству, имеем:

$$(a;\ b) = a + b \cdot i.$$

Принимая во внимание, каким образом были введены действия над комплексными числами, получаем:

$$\boxed{\begin{aligned}&(a + bi) + (c + di) = (a + c) + (b + d)i;\\&(a + bi)(c + di) = (ac - bd) + (ad + bc)i.\end{aligned}}$$

Пример: $(2+i)+(3+4i)=(2+3)+(1+4)i=5+5i$,

$$(2+i)(3+4i)=(2\cdot 3-1\cdot 4)+(2\cdot 4+1\cdot 3)i=2+11i.$$

Этот результат показывает, что желаемое расширение числовой области действительно возможно, и действия над комплексными числами действительно можно производить так, как нам хотелось.

Прежде чем перейти к делению, введем одно новое понятие: понятие *сопряженного числа*.

Комплексное число $\bar{z}=a-bi$ называется сопряженным комплексному числу $z=a+bi$.

Путем несложных вычислений нетрудно убедиться в справедливости следующих соотношений:

$= z$ ($\bar{\bar{z}}$ обозначает число, сопряженное числу $\bar{z}$).
Если z и w — комплексные числа, то $\overline{z+w}=\bar{z}+\bar{w}$ и $\overline{z\cdot w}=\bar{z}\cdot$.
Если $z=a+bi$, то $z+\bar{z}=2a$ — вещественное число,
а $z\cdot\bar{z}=a^2+b^2$ — положительное вещественное число (оно равно 0 только если $z=0$).

Пример: для числа $z=2-3i$ имеем: $\bar{z}=2+3i$, $\bar{\bar{z}}=2-3i$, $z+\bar{z}=4$, $z\cdot\bar{z}=2^2+3^2=13$. Если $w=-5+4i$, то $z+w=-3+i$, $z\cdot w=$ $=2+23i$; поскольку $\bar{w}=-5-4i$, то $\bar{z}+\bar{w}=-3-i=\overline{z+w}$, $\bar{z}\cdot\bar{w}=$ $=2-23i=\overline{z\cdot w}$.

Деление комплексных чисел. Теперь мы покажем, что в области комплексных чисел выполнимо и деление. Рассмотрим сначала случай, когда комплексное число $z=a+bi$ надо разделить на вещественное число N. Очевидно, что при $N\neq 0$ частное от деления есть комплексное число $\frac{a}{N}+\frac{b}{N}i$, так как после умножения на N мы получим z. Предположим теперь, что нужно найти комплексное число x, для которого $wx=z (w\neq 0)$. Если такое число существует, то, умножая последнее равенство на $\bar{w}$, получим: $(\bar{w}w)x=\bar{w}z$. Но поскольку $N=\bar{w}w$ — вещественное число, то обе части равенства можно разделить на N, после чего получаем единственно возможное $x=\frac{\bar{w}z}{N}$. Это число действительно является частным, так как $w\cdot\frac{\bar{w}z}{N}=\frac{w\bar{w}\cdot z}{N}=\frac{Nz}{N}=z$.

Выполним, например, деление $\frac{3+4i}{2+i}$. Рассмотренный выше прием можно выразить словами следующим образом: «числитель и знаменатель следует умножить на число, сопряженное знаменателю». Отсюда: $\frac{3+4i}{2+i}=$ $=\frac{(3+4i)(2-i)}{(2+i)(2-i)}=\frac{10+5i}{3}=2+i$ (действительно, $(2+i)(2+i)=$ $=3+4i$).

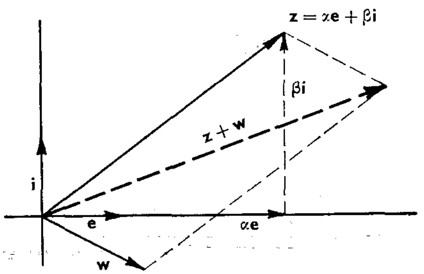
Рассмотрим другой пример: $\frac{5+3i}{3+5i} = \frac{(5+3i)(3-5i)}{(3+5i)(3-5i)} = \frac{30-16i}{34} = \frac{30}{34} - \frac{16}{34}i = \frac{15}{17} - \frac{8}{17}i$.

Расширение понятия числа мы всякий раз производили таким образом, чтобы по возможности «все оставалось в силе». Обычно это оказывается невозможным. Например, комплексные числа уже нельзя упорядочить так же, как вещественные числа. Убедиться в этом нетрудно. Если бы такое упорядочение оказалось возможным, то одно из чисел, i или $-i$, было бы положительным. Тогда, умножая это число само на себя, мы должны были бы получить положительное число. Поскольку $i^2 = (-i)^2 = -1$, то числа -1 и $(-1)^2 = 1$ были бы положительны, но тогда и их сумма $(-1) + 1 = 0$ была бы положительной, что, однако, не имеет места.

ГЕОМЕТРИЧЕСКИЙ СПОСОБ ВВЕДЕНИЯ

Задание комплексных чисел с помощью числовых пар наводит на мысль изображать их в виде точек на плоскости.

Выберем на плоскости прямоугольную систему координат и поставим в соответствие комплексному числу $a + bi$ точку с координатами $(a;\ b)$. Очевидно, что соответствие взаимно однозначно. Следовательно, комплексные числа могут рассматриваться как точки плоскости. Другими словами, плоскость можно рассматривать как модель совокупности комплексных чисел. Изображая комплексные числа в виде точек на плоскости, мы получаем так называемую гауссову числовую плоскость или комплексную плоскость. Для большей геометрической наглядности комплексному числу $a + bi$ можно поставить в соответствие не точку с координатами $(a;\ b)$, а вектор, выходящий из начала координат, с концом в этой точке. Нетрудно видеть, что вектор, соответствующий сумме двух комплексных чисел, оказывается равным сумме векторов, соответствующих слагаемым. Легко убедиться и в том, что длина вектора, соответствующего произведению двух комплексных чисел, равна произведению длин векторов, соответствующих сомножителям. (Ниже это будет доказано.) Эти свойства позволяют создать геометрическую модель комплексных чисел. Начать можно с того, что числовая ось «уже заполнена» вещественными числами, поэтому остальные числа можно «разместить» только на плоскости. Итак, пусть комплексные числа будут векторами на плоскости. Сумму векторов будем определять по правилу сложения векторов, то есть по правилу параллелограмма. Далее, предположим, что произведение векторов есть вектор, длина которого равна произведению длин сомножителей. Поскольку среди комплексных чисел должны быть и все вещественные числа, то предположим, что существует вектор **e** (соответствующий числу 1), при умножении на который все другие векторы не меняются. Рассмотрим



также вектор, перпендикулярный вектору **e** и имеющий ту же длину; обозначим его через **i**. Тогда всякий вектор может быть представлен в виде $\alpha\mathbf{e} + \beta\mathbf{i}$, где α и β — вещественные числа. И, наконец, предположим, что умножение векторов дистрибутивно относительно сложения, откуда следует, что

$$(\alpha\mathbf{e} + \beta\mathbf{i})(\gamma\mathbf{e} + \delta\mathbf{i}) = \alpha\gamma\mathbf{e}^2 + \beta\gamma\mathbf{ie} + \alpha\delta\mathbf{ei} + \beta\delta\mathbf{i}^2.$$

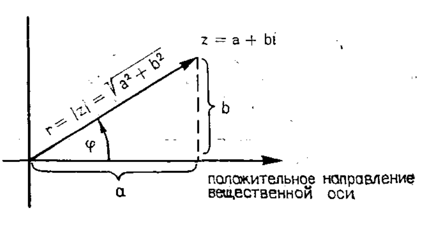
Поскольку $\mathbf{e}^2 = \mathbf{e}$, $\mathbf{ie} = \mathbf{ei} = \mathbf{i}$, то достаточно определить $\mathbf{i}^2$. Об этом векторе известно, что его длина равна квадрату длины вектора **i**. Однако длина вектора **i** равна длине вектора **e**; в свою очередь длина вектора **e** может равняться только 1, так как при умножении на **e** все другие векторы не меняются, и в то же время длина произведения равна произведению длин сомножителей. Следовательно, вектор $\mathbf{i}^2$ имеет длину 1. С другой стороны, и вектор $\mathbf{e} + \mathbf{i}$, и $\mathbf{e} - \mathbf{i}$ имеют длину $\sqrt{2}$, следовательно, их произведение имеет длину 2. Рассмотрим их произведение: $(\mathbf{e} + \mathbf{i})(\mathbf{e} - \mathbf{i}) = \mathbf{e}^2 - \mathbf{ei} + \mathbf{ei} - \mathbf{i}^2 = \mathbf{e} - \mathbf{i}^2 = \mathbf{e} + (-\mathbf{i}^2)$. Из правила сложения векторов следует, что длина суммы векторов не может быть больше суммы длин слагаемых, причем равенство достигается лишь в случае, если слагаемые имеют одинаковое направление. Поскольку векторы **e** и $-\mathbf{i}^2$ имеют длину 1, а их сумма — длину 2, то эти векторы совпадают по направлению, откуда следует, что $-\mathbf{i}^2 = \mathbf{e}$, то есть $\mathbf{i}^2 = -\mathbf{e}$. Но тогда для произведения векторов получается уже известный нам результат.

Комплексные числа находят применение в самых разных областях. Кроме решения уравнений, они могут применяться и в геометрических доказательствах, и, например, в физике для описания колебательных контуров.

ТРИГОНОМЕТРИЧЕСКАЯ И ЭКСПОНЕНЦИАЛЬНАЯ ФОРМЫ КОМПЛЕКСНОГО ЧИСЛА

Рассмотрим вектор, соответствующий комплексному числу $z = a + bi$. Компоненты этого вектора по осям равны, соответственно, a и b. Построив эти отрезки, можно получить некоторый прямоугольный треугольник, катеты которого равны a и b, а гипотенуза имеет длину $\sqrt{a^2 + b^2}$, равную длине рассматриваемого вектора. Число $\sqrt{a^2 + b^2}$ (неотрицательное!) называется модулем комплексного числа z и обозначается через $|z|$.

Угол φ, противоположный стороне b треугольника (см. рис.), называется аргументом комплексного числа z и обозначается через $\arg z$.



Вообще аргументом комплексного числа называется угол, образованный соответствующим вектором с положительным направлением вещественной оси.

Всякое комплексное число имеет однозначно определенный модуль, который равен 0 тогда и только тогда, *когда само комплексное число равно 0; всякое комплексное число, отличное от 0, имеет однозначно определенный аргумент.*

Два комплексных числа равны тогда и только тогда, когда равны их модули и аргументы. В самом деле, если r — модуль комплексного числа z, а φ — его аргумент, то $a = r \cdot \cos\varphi$ и $b = r \cdot \sin\varphi$. Отсюда число определяется однозначно:

$$\boxed{z = r \cdot (\cos\varphi + i \cdot \sin\varphi)\,.}$$

Такой вид называется *тригонометрической формой* комплексного числа z. Число 0 не может быть записано в тригонометрической форме.

Найдем, например, тригонометрическую форму числа $1 + i$. Модуль этого числа есть $\sqrt{1+1} = \sqrt{2}$. Далее, $\cos\varphi = \dfrac{1}{\sqrt{2}}$, то есть для φ возможны два значения: $45°$ и $360°—45°$. Поскольку в нашем случае $\sin\varphi$ положителен, то подходит лишь $\varphi = 45°$. Таким образом, $1 + i = 2 \cdot (\cos 45° + i \cdot \sin 45°)$.

Тригонометрическая форма оказывается особенно удобной для умножения комплексных чисел. Пусть

$$w = \varrho(\cos\psi + i \cdot \sin\psi)\,.$$

Вычислим произведение zw:

$$zw = r\varrho(\cos\varphi + i \cdot \sin\varphi)(\cos\psi + i \cdot \sin\psi) = r\varrho(\cos\varphi\cos\psi -$$
$$- \sin\varphi \cdot \sin\psi + i \cdot \cos\varphi \cdot \sin\psi + i \cdot \sin\varphi \cdot \cos\psi) = r\varrho\big(\cos(\varphi+\psi) +$$
$$+ i \cdot \sin(\varphi+\psi)\big)\,.$$

Это так называемая формула Муавра, согласно которой модуль произведения комплексных чисел равен произведению модулей, а его аргумент равен сумме аргументов сомножителей.

Применение формулы Муавра значительно упрощает деление, возведение в степень и извлечение корня из комплексных чисел. Для деления получаем:

$$\frac{z}{w} = \frac{r}{\varrho}\big(\cos(\varphi-\psi) + i \cdot \sin(\varphi-\psi)\big)\,.$$

Для возведения в степень:

$$z^n = r^n\big(\cos(n\varphi) + i \cdot \sin(n\varphi)\big)\,.$$

Извлечение корня из комплексного числа — действие не однозначное. Точнее: корень n-ой степени из комплексного числа, отличного от 0, имеет n различных значений, которые могут быть найдены по следующей формуле:

$$\boxed{\sqrt[n]{z}=\sqrt[n]{r}\cdot\left(\cos\frac{\varphi+2k\pi}{n}+i\cdot\sin\frac{\varphi+2k\pi}{n}\right),\text{ где } k=1,2,\ldots,n.}$$

Рассмотрим несколько примеров выполнения действий над комплексными числами, заданными в тригонометрической форме. Пусть $z=1+i, w=-1+\sqrt{3}\cdot i$. Как мы уже видели, $z=\sqrt{2}\cdot(\cos 45^\circ+i\cdot\sin 45^\circ)$; $|w|=\sqrt{1+3}=\sqrt{4}=2$. Если $\arg w=\varphi$, то $\cos\varphi=-\frac{1}{2}$, то есть φ имеет одно из двух значений: $180^\circ-60^\circ=120^\circ$ или $180^\circ+60^\circ$. Поскольку $\sin\varphi$ положителен, то $\varphi=120^\circ$, то есть $w=2\cdot(\cos 120^\circ+i\cdot\sin 120^\circ)$. Отсюда:

$$zw=2\cdot\sqrt{2}\cdot(\cos 165^\circ+i\cdot\sin 165^\circ)\ \text{ и }\ \frac{w}{z}=\sqrt{2}\cdot(\cos 75^\circ+i\cdot\sin 75^\circ)$$

или, например,

$$z^{12}=(\sqrt{2})^{12}\cdot(\cos 540^\circ+i\cdot\sin 540^\circ)=2^6(\cos 180^\circ+i\cdot\sin 180^\circ)=$$
$$=64\cdot(-1+i\cdot 0)=-64.$$

Найдем, например, значения $\sqrt[7]{1}$ (приближенно), то есть определим корни седьмой степени из единицы (вообще значения корня $\sqrt[n]{1}$ называют *корнями n-ой степени из единицы*). Имеем: $1=1\cdot(\cos 0^\circ+i\cdot\sin 0^\circ)$. $\sqrt[7]{1}=1$ *(в области вещественных чисел извлечение корня — однозначное действие)*. Вычислим необходимые углы: $\frac{360^\circ}{7}=51^\circ 26'$; $\frac{720^\circ}{7}\,102^\circ 51'$; $\frac{1080^\circ}{7}=154^\circ 17'$; $\frac{1440^\circ}{7}=205^\circ 43'$; $\frac{1800^\circ}{7}=257^\circ 9'$; $\frac{2160^\circ}{7}=308^\circ 34'$; (и 0°). Разыскав по таблицам тригонометрических функций соответствующие значения синуса и косинуса, получаем следующие значения корней: $0{,}6235+0{,}7818\cdot i$; $-0{,}2225+0{,}9749\cdot i$; $-0{,}9010+0{,}4339\cdot i$; $-0{,}9010-0{,}4339\cdot i$; $-0{,}2225-0{,}9749\cdot i$; $0{,}6235-0{,}7818\cdot i$ и 1.

С помощью возведения комплексных чисел в степень нетрудно установить некоторые тригонометрические соотношения. Выразим, например, $\cos(4\varphi)$ и $\sin(4\varphi)$ через $\cos\varphi$ и $\sin\varphi$:

$$(\cos(4\varphi)+i\cdot\sin(4\varphi))=(\cos\varphi+i\cdot\sin\varphi)^4=\cos^4\varphi+4i\cdot\cos^3\varphi\cdot\sin\varphi+$$
$$+(-6)\cos^2\varphi\cdot\sin^2\varphi+4(-i)\cos\varphi\cdot\sin^3\varphi+\sin^4\varphi.$$

Отсюда:

$$\boxed{\begin{aligned}\cos(4\varphi) &= \cos^4\varphi - 6\cos^2\varphi\cdot\sin^2\varphi + \sin^4\varphi,\\ \sin(4\varphi) &= 4\cos^3\varphi\cdot\sin\varphi - 4\cos\varphi\cdot\sin^3\varphi.\end{aligned}}$$

В заключение найдем тригонометрическую форму числа $\bar{z}$. Очевидно, что $\bar{z} = r(\cos\varphi - i\cdot\sin\varphi)$, только это еще не тригонометрическая форма, так как перед i стоит минус, а должен стоять плюс. Посмотрим, что означает $\bar{z}$ с геометрической точки зрения. Легко видеть, что $\bar{z}$ получается в результате зеркального отображения числа z относительно вещественной оси. Отсюда следует, что модули этих чисел равны, в то время как $\arg\bar{z} = -\arg z$. Таким образом, приходим к следующей записи:

$$\boxed{\bar{z} = r\big(\cos(-\varphi) + i\cdot\sin(-\varphi)\big).}$$

Это выражение, естественно, равно предыдущему, так как $\cos(-\varphi) = \cos\varphi$ и $\sin(-\varphi) = -\sin\varphi$.

Аргумент комплексного числа ведет себя так же, как показатель степени. Это обстоятельство наводит на мысль представлять комплексные числа в следующем виде:

$$\boxed{z = r\cdot e^{i\varphi},}$$

где буква e обозначает основание натурального логарифма (см. гл. «Математический анализ»). Можно показать, что к комплексным числам, записанным в таком виде, формально применимы правила действий над степенью. Запись $z = r\cdot e^{i\varphi}$ называют *экспоненциальной формой* комплексного числа. Представляя числа z и $\bar{z}$ в экспоненциальной форме, получаем:

$$\cos\varphi + i\cdot\sin\varphi = e^{i\varphi};\quad \cos\varphi - i\cdot\sin\varphi = e^{-i\varphi}.$$

Выражая отсюда $\cos\varphi$ и $\sin\varphi$, приходим к так называемым *формулам Эйлера:*

$$\boxed{\cos\varphi = \frac{e^{i\varphi} + e^{-i\varphi}}{2};\quad \sin\varphi = \frac{e^{i\varphi} - e^{-i\varphi}}{2i}.}$$

Модуль комплексного числа z также может быть записан в виде степени числа e. Пусть $r = e^u$, тогда

$$z = e^{u+i\varphi}.$$

Такое представление дает возможность определить логарифм комплексного числа.

ДАЛЬНЕЙШЕЕ ОБОБЩЕНИЕ ПОНЯТИЯ ЧИСЛА

До сих пор по мере обобщения понятия числа мы познакомились с тремя важными числовыми множествами: с рациональными, вещественными и комплексными числами.

Для всех этих числовых множеств характерно то, что в них однозначно выполнимы четыре арифметических действия: сложение, вычитание, умножение и деление (деление на 0 мы исключили из рассмотрения). Кроме того, были выполнены свойства действий: коммутативность и ассоциативность сложения и умножения, а также дистрибутивность умножения относительно сложения. Числовые системы, в которых выполнены все перечисленные свойства, называются *числовыми полями*. Таким образом, говорят о *поле рациональных чисел*, *поле вещественных чисел* и о *поле комплексных чисел*.

В поле комплексных чисел мы имели две так называемые «единицы»: числа 1 и i, с помощью которых любое комплексное число можно однозначно представить в виде $a \cdot 1 + b \cdot i$. Возникла, однако, необходимость и во введении так называемых *гиперкомплексных* чисел, определяемых несколькими «единицами».

КВАТЕРНИОНЫ

Попытка ввести систему чисел, содержащую три единицы, потерпела неудачу. Однако построение системы чисел с четырьмя единицами по существу удалось.

В этом случае получается так называемая система кватернионов, то есть «чисел» вида

$$A = a + bi + cj + dk,$$

где a, b, c и d — вещественные числа.

Если $B = x + yi + uj + vk$ — некоторый другой кватернион, то равенство $A = B$ имеет место тогда и только тогда, когда $a = x$, $b = y$, $c = u$, $d = v$.

Действия над кватернионами. Сумма и произведение кватернионов определяются следующим образом:

$$A + B = (a + x) + (b + y)i + (c + u)j + (d + v)k,$$

$$AB = (ax - by - cu - dv) + (ay + bx + cv - du)i +$$
$$+ (au - bv + cx + dy)j + (av + bu - cy + dx)k.$$

Очевидно, что так определенное сложение ассоциативно и коммутативно. Нетрудно убедиться и в ассоциативности умножения, однако коммутативность умножения уже не имеет места. Так, например, $ij = k$, но $ji = -k$. Это обстоятельство вызывает некоторые трудности в связи с дистрибутивностью. Дело в том, что здесь приходится рассматривать два типа дистрибутивности: $(A + B)C = AC + BC$ и $C(A + B) = CA + CB$. Если бы умножение было коммутативно, то справедливость одного из записанных равенств влекла бы за собой справедливость другого. В данном же случае выполнение этих равенств приходится доказывать отдельно. Тем не менее *можно показать, что оба тождества действительно имеют место.*

Вычитание не вызывает особых трудностей, однако выполнимость деления далеко не очевидна. Ясно лишь, что деление кватерниона на вещественное число можно выполнять «почленно». Для рассмотрения деления в общем случае здесь также целесообразно ввести понятие «сопряженного кватерниона».

Кватернионом, сопряженным с A является кватернион $\bar{A} = a - bi - cj - dk$. Сопряженные кватернионы обладают следующими свойствами (аналогичными свойствам сопряженных комплексных чисел):

$\overline{A+B} = \bar{A} + \bar{B}$, $\overline{AB} = \bar{B} \cdot \bar{A}$ (это не ошибка, порядок сомножителей действительно меняется!)
$A + \bar{A} = 2a$ — вещественное число, $A\bar{A} = a^2 + b^2 + c^2 + d^2$ — неотрицательное вещественное число.

Последнее соотношение позволяет свести деление кватернионов к делению кватерниона на вещественное число. Действительно, если $AX = B$, то, умножая обе части этого равенства (слева!) на $\bar{A}$, получаем: $(\bar{A}A)X = \bar{A}B$. Отсюда уже можно найти X. Аналогичным образом можно говорить и о делении «справа», то есть о решении уравнения $YA = B$. В этом случае после умножения на $\bar{A}$ справа получается уравнение $Y(A\bar{A}) = B\bar{A}$, которое также разрешимо. Если бы умножение было коммутативным, то кватернионы образовывали бы числовое поле. Тем не менее, систему кватернионов все же называют *полем кватернионов* относительно действий, определенных выше.

Тот факт, что в поле кватернионов не выполняется коммутативность умножения, приводит к многочисленным курьёзам. При рассмотрении многочленов (см. стр. 71) мы увидим, что всякий многочлен n-ой степени имеет не более n корней (в поле комплексных чисел). В поле кватернионов это не так! Можно, например, сразу указать *шесть* корней многочлена $x^2 + 1$: $i, -i, j, -j, k, -k$. Больше того, этот многочлен имеет в поле кватернионов бесконечно много корней! Все корни могут быть получены по следующей формуле: $X = pi + qj + rk$, где $p^2 + q^2 + r^2 = 1$. Вообще, если кватерниону $pi + qj + rk$ поставить в соответствие точку трехмерного пространства с координатами $(p; q; r)$, то решениям уравнения будут соответствовать точки сферы радиуса 1.

Геометрическое представление. Оказывается, что между кватернионами и точками — а точнее векторами — пространства существует значительно более тесная связь. (Кстати, Гамильтон ввел кватернионы, чтобы выполнять деление векторов).

Предположим, что в кватернионах A и B $a = x = 0$, то есть

$$A = bi + cj + dk, \quad B = yi + uj + vk.$$

Такие кватернионы называют чистыми кватернионами. Всякому чистому кватерниону можно поставить в соответствие единственный вектор пространства (и наоборот). Таким образом, кватерниону A соответствует вектор с компонентами (b; c; d), а кватерниону B—вектор с компонентами (y; u; v). В этом случае можно показать, что $-\frac{AB+BA}{2}$ равняется скалярному произведению этих векторов, а вектор, соответствующий кватерниону $\frac{AB-BA}{2}$ (так как это тоже чистый кватернион) представляет собой их векторное произведение.

Теорема Фробениуса. Возникает вопрос, возможно ли расширение поля вещественных чисел, аналогичное изложенному выше. Сформулируем несколько точнее, что понимается в данном случае под «аналогией».

Некоторое «числовое поле» называется расширением (конечного ранга) поля вещественных чисел, если

1. в нем однозначно выполнимы четыре арифметических действия и имеют место (хотя бы частично) свойства действий;

2. оно содержит поле вещественных чисел;

3. оно содержит конечное число фиксированных «чисел», обладающих тем свойством, что любое «число» поля может быть единственным образом представлено в виде их суммы с вещественными коэффициентами; если можно указать n таких чисел, то говорят о расширении ранга n.

Имеет место следующая *теорема Фробениуса:* существуют лишь два расширения конечного ранга поля вещественных чисел, в которых выполнены все свойства действий: поле вещественных чисел (с рангом 1) и поле комплексных чисел (с рангом 2). Существуют только три расширения (конечного ранга) поля вещественных чисел, в которых выполнены все свойства действий, за исключением коммутативности умножения: поле вещественных чисел, поле комплексных чисел и поле кватернионов (с рангом 4).

Дальнейшие возможности. Следует упомянуть и дальнейшие результаты. Как в поле комплексных чисел, так и в поле кватернионов мы производили деление с помощью умножения на сопряженный множитель и при этом пользовались ассоциативностью умножения. Однако ассоциативность нужна была лишь в тех случаях, когда сомножители были сопряженными друг другу. Этот случай можно свести к случаю, когда оба сомножителя равны между собой. Если вместо ассоциативности потребовать лишь выполнения так называемого *альтернативного тождества:* $(AA)B = A(AB)$ и $A(BB) = (AB)B$, то можно указать еще одно расширение — так называемую *алгебру Кэли—Диксона,* ранг которой равен 8. (О ранге будет сказано подробнее в главе «Линейная алгебра».)

ДРУГАЯ ВОЗМОЖНОСТЬ ОБОБЩЕНИЯ: p-АДИЧЕСКИЕ ЧИСЛА

Йожеф Кюршак ввел так называемые *p-адические числа,* которые играют важную роль в алгебре. Дадим сначала определение *целых p-адических чисел.*

Пусть p — некоторое фиксированное простое число, и пусть $a_0, a_1, \ldots, a_n, \ldots$ неотрицательные целые числа, меньшие p. Формально составленное выражение

$$a_0 + a_1 p + a_2 p^2 + \ldots + a_n p^n + \ldots$$

называется целым p-адическим числом. Если $a_0 = 1$ и $a_1 = a_2 = \ldots = 0$, то получим *p-адическую единицу*, при $a_0 = a_1 = \ldots = 0$ — *p-адический нуль*.

Два p-адических числа считаются равными, если равны их «соответственные коэффициенты». Целые p-адические числа можно рассматривать как «бесконечные» числа, записанные в p-ичной системе счисления (см. гл. «Теория чисел»). Действия над ними производятся точно так же, как и над числами, записанными в p-ичной системе, однако, здесь сложение и умножение «никогда не заканчиваются». Если же, начиная с некоторого индекса, все числа a_i равны 0, то мы получим конечные суммы. В этом случае запись этих чисел тождественна записи чисел в p-ичной системе счисления, и действия над ними производятся совершенно аналогично действиям над p-ичными числами. То обстоятельство, что действия никогда не заканчиваются, не является принципиальной проблемой, так как, например, и действия над бесконечными десятичными дробями, вообще говоря, нельзя произвести за конечное число шагов.

Приводим несколько примеров. Пусть $p = 5$ и рассмотрим следующие 5-адические числа:

$$A = 2 + 3 \cdot 5 + 1 \cdot 5^2 + 3 \cdot 5^3 + 1 \cdot 5^4 + 3 \cdot 5^5 + \ldots$$

$$B = 2 + 4 \cdot 5 + 3 \cdot 5^2 + 2 \cdot 5^3 + 4 \cdot 5^4 + 3 \cdot 5^5 + \ldots$$

В числе A повторяется группа цифр (3, 1), а в числе B — группа (2, 4, 3). Действия над бесконечными десятичными дробями производились следующим образом: мы последовательно брали все более точные десятичные приближения исходных дробей и над полученными конечными дробями выполняли требуемые действия, получаемые при этом результаты всякий раз давали более точное приближение искомого результата. Аналогичным образом мы поступим и в данном случае. Будем рассматривать последовательно 3, 6 и так далее слагаемых каждого числа и производить соответствующее действие в пятеричной системе счисления, тем самым мы каждый раз будем получать все больше знаков искомого результата. (Можно доказать, что «первые» 3, 6 и так далее знаков результата будут «точными». Дело в том, что здесь, в отличие от бесконечных десятичных дробей, отброшенный «знак» не влияет на предыдущий. Например, при вычислении суммы $0{,}38\ldots + 0{,}25\ldots$ с точностью до одного десятичного знака получаем: $0{,}3 + 0{,}2 = 0{,}5$; если же рассмотреть два знака, то $0{,}38 + 0{,}25 = 0{,}63$, то есть первый знак после запятой изменился). Сложение чисел в пятеричной системе счисления дает:

$$\begin{array}{r} 132 \\ 342 \\ \hline 1024 \end{array} \qquad \begin{array}{r} 313132 \\ 342342 \\ \hline 1211024 \end{array} \qquad \begin{array}{r} 131313132 \\ 342342342 \\ \hline 1024211024 \end{array}$$

Запишем полученные результаты один под другим:

$$\begin{array}{r} 1\,024 \\ 1\,211\,024 \\ 1\,024\,211\,024 \end{array}$$

Отсюда видно, что в сумме будет повторяться группа цифр (4, 2, 0, 1, 1, 2). Таким образом, сумма имеет следующий вид:

$$A + B = 4 + 2 \cdot 5 + 0 \cdot 5^2 + 1 \cdot 5^3 + 1 \cdot 5^4 + 2 \cdot 5^5 + \ldots$$

Вычисляя аналогичным образом произведение AB чисел A и B, получаем, что и здесь будет повторяться группа из шести цифр:

$$AB = 4 + 4 \cdot 5 + 2 \cdot 5^2 + 2 \cdot 5^3 + 1 \cdot 5^4 + 1 \cdot 5^5 + \ldots$$

Можно показать также, что $3A = 1$, а $124B = -97 \cdot 1$, где 3 и -97 можно идентифицировать с целыми p-адическими числами $3 + 0 \cdot p + 0 \cdot p^2 + \ldots$ $-97 + 0 \cdot p + 0 \cdot p^2 + \ldots$

Вообще всякая дробь, знаменатель которой не является кратным p, может быть записана в виде «периодического» p-адического числа (верно и обратное утверждение). Таким образом, в области целых p-адических чисел можно выполнять деление на целые числа, взаимно простые с числом p. Деление на любые целые числа, кроме 0, выполнимо уже в области всех p-адических чисел, которые в общем виде записываются следующим образом:

$$a_{-k}p^{-k} + \ldots + a_{-1}p^{-1} + a_0 + a_1 p + a_2 p^2 + \ldots + a_n p^n + \ldots$$

Больше того, можно показать, что в области всех p-адических чисел деление вообще выполнимо. Отсюда следует, что мы получили некоторое «числовое поле». Общий вид p-адического числа поразительно похож на разложение некоторой функции в так называемый ряд Лорана. Это может быть использовано для образования *формальных степенных рядов*. Так называют *формально* составленные *выражения* вида

$$a_{-k}x^{-k} + \ldots + a_{-1}x^{-1} + a_0 + a_1 x + a_2 x^2 + \ldots + a_n x^n + \ldots$$

где коэффициенты a_i — произвольные вещественные или комплексные числа. Сложение таких рядов выполняется почленно, а их произведение можно получить как *произведение Коши* (см. гл. «Математический анализ»), при этом вопрос о сходимости не принимается во внимание. О сходимости здесь и не имеет смысла говорить, так как вместо x мы *никогда не подставляем никакого конкретного значения*. В области формальных степенных рядов однозначно выполнимы четыре основные действия, и справедливы свойства действий. Формальные степенные ряды могут быть с успехом использованы, например, при исследовании производящих функций, так как и там не совершается подстановка.

МНОГОЧЛЕНЫ

КОЛЬЦО МНОГОЧЛЕНОВ

Основной задачей классической алгебры было решение (алгебраических) уравнений. Под уравнением понимали то, что получается, когда некоторое выражение — в котором могут присутствовать известные и неизвестные величины, а также их сумма, разность и произведение — равняется нулю. При решении уравнения производили всевозможные преобразования этого выражения. Именно поэтому еще до решения уравнения целесообразно посмотреть, какие преобразования и действия можно производить с данным выражением. Для начала мы будем рассматривать лишь такие выражения, в которых присутствует лишь одно неизвестное. Поскольку речь пока будет идти не об уравнениях, то неизвестное не следует рассматривать как число, а лишь как неопределенную величину, над которой можно производить действия точно так же, как над числами. Если в некотором выражении присутствует только одно неизвестное, то выражение можно записать в виде суммы степеней неизвестного, умноженных на некоторые коэффициенты. Для решения уравнения важно знать также, какие числа присутствуют в уравнении, и каким числом является само решение уравнения, то есть являются ли они, например, рациональными, вещественными или комплексными числами. В каждом из этих случаев мы имеем дело с числами из некоторого числового поля. Исходя из этих соображений, дадим теперь следующее определение:

пусть $a_0, a_1, \ldots, a_n$ — элементы некоторого числового поля T. Выражение

$$f(x) = a_0 + a_1x + a_2x^2 + \ldots + a_nx^n$$

называется многочленом (или полиномом) над полем T, при этом числа $a_0, a_1, \ldots, a_n$ называются коэффициентами многочлена.

Чтобы производить операции над многочленами, надо сначала определить, когда два многочлена равны между собой.

Многочлены $f(x)$ и $g(x) = b_0 + b_1x + b_2x^2 + \ldots + b_kx^k$ равны тогда и только тогда, когда $a_0 = b_0$, $a_1 = b_1, \ldots$

Прежде чем перейти к определению операций над многочленами, введем несколько понятий.

Если $a_n \neq 0$, то многочлен $f(x)$ называется многочленом степени n. В этом случае используется обозначение $cm(f) = n$.*

Если все коэффициенты многочлена $f(x)$ равны 0, то $f(x)$ не имеет степени.

Если $a_n \neq 0$, то a_n называется старшим коэффициентом многочлена; a_0 называется постоянным членом многочлена.

Из определения равенства многочленов следует, что при $k < n$ многочлен $g(x)$ может быть записан в виде

$$g(x) = b_0 + b_1x + b_2x^2 + \ldots + b_kx^k + b_{k+1}x^{k+1} + \ldots + b_nx^n,$$

где $b_{k+1} = \ldots = b_n = 0$.

* В оригинале используется символ «gr» — сокращение латинского слова gradus.

ОПЕРАЦИИ НАД МНОГОЧЛЕНАМИ

Сложение и умножение. Сумма и произведение двух многочленов определяются следующим образом:

$$f(x) + g(x) = (a_0 + b_0) + (a_1 + b_1)x + (a_2 + b_2)x^2 + \ldots + (a_n + b_n)x \ ;$$

$$f(x)\,g(x) = a_0b_0 + (a_0b_1 + a_1b_0)x + (a_0b_2 + a_1b_1 + a_2b_0)x^2 + \ldots + a_nb^kx^{n+k}.$$

Нетрудно видеть, что сложение и умножение коммутативны и ассоциативны. Можно доказать и выполнение дистрибутивности умножения относительно сложения.

Вычитание. Можно показать, что в области многочленов однозначно выполнимо и вычитание:

$$\boxed{f(x) - g(x) = (a_0 - b_0) + (a_1 - b_2)x + (a_2 - b_2)x^2 + \ldots + (a_n - b_n)x^n.}$$

Если это не приводит к недоразумению, то мы будем обозначать многочлены $f(x)$ и $g(x)$, соответственно, через f и g. Точно так же вместо $f(x) + g(x)$ и $f(x)\,g(x)$ будем писать, соответственно, $f + g$ и fg. (Легко видеть, что сумма, разность и произведение многочленов над полем T также оказываются многочленами над полем T. Аналогичным свойством обладают и целые числа. Структуры подобного «типа» встречаются довольно часто, поэтому они получили особое название. Здесь мы дадим название лишь системе многочленов над полем T, однако, позже вернемся к этому понятию.)

Систему многочленов над полем T называют кольцом многочленов над полем T и обозначают через $T[x]$.

Здесь T указывает на числовое поле, откуда берутся коэффициенты многочленов, а x обозначает «неизвестное».

Зная степени многочленов f и g, нетрудно определить степень многочлена fg: очевидно, что $cm(fg) = cm(f) + cm(g)$.

О степени суммы $f + g$ заранее можно сказать лишь, что она не может превысить ни степень f, ни степень g. Обычно $cm(f + g)$ совпадает с наибольшим из чисел $cm(f)$ и $cm(g)$, однако, если $cm(f) = cm(g)$, то может случиться, что степень $f + g$ окажется меньше.

Пример: если $f(x) = 2 + 3x + 5x^3$, $g(x) = 3 + x^2 - 5x^3$, то $f(x) + g(x) = 5 + 3x + x^2$. Здесь $cm(f + g) = 2$, хотя $cm(f) = cm(g) = 3$. То же самое справедливо и для разности двух многочленов.

Деление. Точно так же, как в области целых чисел, в системе многочленов деление, вообще говоря, не выполнимо. Для целых чисел деление в некоторой степени удалось заменить делением с остатком (см. гл. «Теория чисел», стр. 164). Деление с остатком можно определить и для многочленов. При делении целых чисел остаток всегда оказывался меньше делителя. Однако в системе многочленов не может идти речь о том, что один многочлен меньше другого. Здесь эту функцию выполняет степень многочленов. Нетрудно показать, что в системе многочленов можно производить деление с остатком таким образом, чтобы степень остатка оказывалась меньше степени делителя. Точнее:

если $f(x)$ — произвольный многочлен, а $g(x)$ — многочлен, отличный от 0, то найдутся такие многочлены $q(x)$ и $r(x)$, что

$$f(x) = q(x)g(x) + r(x), \qquad \text{причем или } cm(r) < cm(g), \text{ или } r(x) = 0.$$

Важно отметить, что «$r(x) = 0$» не является уравнением, а означает лишь, что все коэффициенты многочлена $r(x)$ равны 0! Случай $g(x) = 0$ приходится исключить из рассмотрения, так как степень «нулевого» многочлена не определена.

Если $cm(f) < cm(g)$, то, полагая $q(x) = 0$ и $r(x) = f(x)$, получаем: $f(x) = 0\ g(x) + f(x)$, что уже имеет указанный вид. Если же $cm(f) \geq cm(g)$, то пусть $f(x) = a_nx^n + \ldots + a_1x + a_0$ и $g(x) = b_kx^k + \ldots + b_1x + b_0$, где $n \geq k$, $a_n \neq 0$, $b_k \neq 0$ (при делении многочлены удобнее записывать в порядке убывания степеней неизвестного).

Имеем: $\frac{a_n}{b_k} \cdot x^{n-k} \cdot g(x) = a_nx^n + \ldots + \frac{b_0a_n}{b_k} x^{n-k}$. Таким образом степень многочлена $f_1(x) = f(x) - \frac{a_n}{b_k} \cdot x^{n-k} \cdot g(x)$ оказывается меньше n. Продолжая такой процесс дальше, мы получим остаток, степень которого меньше k. Этот остаток и будет многочленом $r(x)$. Многочлен $q(x)$ будет иметь вид $\frac{a_n}{b_k} x^{n-k} + \ldots$

Пример: Для простоты рассмотрим многочлены с целыми коэффициентами. Кроме того, во избежание дробных коэффициентов, выберем старший коэффициент делителя равным 1. Пусть $f(x) = 3x^5 - 2x^4 + 7x^2 - x + 11$ и $g(x) = x^2 - x + 3$. Следуя указанному приему, рассмотрим многочлен $f_1(x) = f(x) - 3x^{5-2} \cdot g(x)$. Таким образом из $f(x)$ нужно вычесть многочлен $3x^3 \cdot g(x) = 3x^5 - 3x^4 + 9x^3$. Точно так же, как и для чисел, вычитание многочленов удобнее производить, записывая их один под другим. При этом следует учесть, что в многочлене $f(x)$, например, отсутствует член третьей степени. Чтобы случайно не вычесть член третьей степени из члена второй степени, имеет смысл выписать и этот член — с коэффициентом 0:

$$\begin{array}{l} 3x^5 - 2x^4 + 0x^3 + 7x^2 - x + 11 \\ -\ (3x^5 - 3x^4 + 9x^3) \end{array}$$

При конкретных вычислениях нередко бывает трудно следить за тем, с каким знаком следует взять тот или иной член, поэтому проще оказывается сразу поменять все знаки в вычитаемом, в результате операция сведется к сложению соответственных членов (разумеется, при сложении знаки следует принимать во внимание). Обычно не имеет смысла заново переписывать многочлены, достаточно лишь изменить знаки вычитаемого на противоположные, записав новые знаки под старыми. После этого при сложении следует всякий раз принимать во внимание «нижний знак»:

$$\begin{array}{r} 3x^5 - 2x^4 + 0x^3 + 7x^2 - x + 11 \\ \pm 3x^5 \mp 3x^4 \pm 9x^3 \qquad\qquad\qquad \\ \hline 0 \qquad 1x^4 - 9x^3 + 7x^2 - x + 11 \end{array}$$

После выполнения первого шага нужно составить многочлен $f_2(x) = f_1(x) - x^2 \cdot g(x)$ и так далее. Такой процесс деления многочленов, аналогично делению чисел, можно изобразить следующим образом:

$$3x^5 - 2x^4 + 0x^3 + 7x^2 - x + 11 : (x^2 - x + 3) = 3x^3 + x^2 - 8x - 4$$

$$\begin{array}{l} \pm 3x^5 \mp 3x^4 \pm 9x^3 \\ \hline \qquad + x^4 - 9x^3 + 7x^2 \\ \qquad \pm x^4 \mp x^3 \pm 3x^2 \\ \hline \qquad\qquad - 8x^3 + 4x^2 - x \\ \qquad\qquad \mp 8x^3 \pm 8x^2 \mp 24x \\ \hline \qquad\qquad\qquad - 4x^2 + 23x + 11 \\ \qquad\qquad\qquad \mp 4x^2 \pm 4x \mp 12 \\ \hline \qquad\qquad\qquad\qquad + 19x + 23 \end{array}$$

Таким образом $f(x) = g(x) \cdot (3x^3 + x^2 - 8x - 4) + (19x + 23)$.

Рассмотрим теперь пример, когда старший коэффициент делителя не равен 1. Пусть $f(x) = x^5 + 3x^4 - x^3 - x^2 + 8x + 37$, а $g(x) = 2x^3 + x^2 - 2x + 5$. И в этом случае можно избежать появления дробных коэффициентов. Для этого будем делить на $g(x)$ не $f(x)$, а многочлен $8 \cdot f(x)$ (вообще берется многочлен $b_k^{n-k+1} \cdot f(x)$):

$$8x^5 + 24x^4 - 8x^3 - 8x^2 + 64x + 296 : (2x^3 + x^2 - 2x + 5) = $$
$$= 4x^2 + 10x - 5$$

$$\begin{array}{l} \pm 8x^5 \pm 4x^4 \mp 8x^3 \pm 20x^2 \\ \hline \qquad + 20x^4 \qquad\quad - 28x^2 + 64x \\ \qquad \pm 20x^4 \pm 10x^3 \mp 20x^2 \pm 50x \\ \hline \qquad\qquad - 10x^3 - 8x^2 + 14x + 296 \\ \qquad\qquad \mp 10x^3 \mp 5x^2 \pm 10x \mp 25 \\ \hline \qquad\qquad\qquad - 3x^2 + 4x + 321 \end{array}$$

Таким образом $8 \cdot f(x) = g(x) \cdot (4x^2 + 10x - 5) + (-3x^2 + 4x + 321)$, откуда после деления на 8 получаем:

$$f(x) = g(x) \cdot \left(\frac{1}{2}x^2 + \frac{5}{4}x - \frac{5}{8}\right) + \left(-\frac{3}{8}x^2 + \frac{1}{2}x + \frac{321}{8}\right).$$

ДЕЛИМОСТЬ МНОГОЧЛЕНОВ; НЕПРИВОДИМЫЕ МНОГОЧЛЕНЫ

Аналогично делимости целых чисел можно говорить и о делимости (нацело) многочленов.

Многочлен $g(x)$ называется делителем многочлена $f(x)$ (обозначается как $g(x) \mid f(x)$ или просто $g \mid f$), если существует многочлен $h(x)$ такой, что $f(x) = h(x)g(x)$. Из рассмотренного выше процесса деления с остатком следует, что если $f(x)$ и $g(x)$ — многочлены над полем T, то $q(x)$ и $r(x)$ также оказываются многочленами над полем T. Однако далеко не ясно, что если $f(x)$ и $g(x)$ — многочлены над полем T и $f(x) = h(x)g(x)$, то и $h(x)$ есть многочлен над полем T. Тем не менее и это может быть показано с помощью того же процесса деления. Пусть $f(x)$ и $g(x)$ — многочлены над полем T, для которых справедливо соотношение $f(x) = h(x)g(x)$. Если $g(x) \neq 0$, то можно выполнить деление с остатком, откуда следует, что над полем T существуют такие многочлены $q(x)$ и $r(x)$, что имеет место равенство $f(x) = q(x)g(x) + r(x)$. Из двух выражений для $f(x)$ получаем:

$$h(x)g(x) = q(x)g(x) + r(x), \text{ то есть } [h(x) - q(x)]g(x) = r(x).$$

Если бы сомножитель $h(x) - q(x)$ был отличен от 0, то степень произведения в левой части последнего равенства была бы равна сумме степеней сомножителей, то есть была бы не меньше степени $g(x)$. Однако это невозможно, так как степень многочлена $r(x)$, стоящего справа, по условию меньше степени $g(x)$. Следовательно, $h(x) - q(x) = 0$, но тогда $h(x) = q(x)$, а значит $r(x) = 0$. Поскольку $q(x)$ — многочлен над полем T, и $h(x) = q(x)$, то и $h(x)$ — многочлен над полем T.

Заметим, что с помощью аналогичных рассуждений можно доказать и однозначность деления с остатком, то есть тот факт, что многочлены $q(x)$ и $r(x)$ определяются однозначно.

Поскольку определена делимость (нацело) многочленов, то можно говорить и о *наибольшем общем делителе* двух или нескольких многочленов.

Многочлен $d(x)$ называется наибольшим общим делителем многочленов $f(x)$ и $g(x)$ и обозначается символом $d = (f, g)$, если $d(x)$ является делителем многочленов $f(x)$ и $g(x)$ и делится (нацело) на любой общий делитель этих многочленов.

Можно показать, что любые два многочлена над полем T имеют наибольший общий делитель, также являющийся многочленом над полем T. Доказательство проводится аналогично доказательству соответствующего утверждения для целых чисел. Если рассмотреть совокупность многочленов вида $f(x)u(x) + g(x)v(x)$, где $u(x)$ и $v(x)$ — произвольные многочлены над полем T, то можно показать, что многочлен наименьшей степени из этой совокупности является наибольшим делителем многочленов $f(x)$ и $g(x)$. Однако этот многочлен определен не однозначно: его произведение на любое число из поля T также является наибольшим общим делителем рассматриваемых многочленов. Вследствие этого принято считать, что старший коэффициент наибольшего общего делителя равен 1. Это уже обеспечивает однозначность определения.

В области целых чисел с помощью наибольшего общего делителя можно было доказать единственность разложения числа на простые множители. Нечто аналогичное имеет место и для многочленов. Роль простых множителей здесь выполняют *неприводимые многочлены*.

Многочлен $f(x)$ над полем T называется неприводимым в поле T, если он не может быть представлен в виде произведения двух многочленов меньшей степени с коэффициентами из поля T.

Очень важно при этом, над каким именно полем рассматривается многочлен. Например, многочлен $x^2 - 2$ неприводим в поле рациональных чисел, однако в поле вещественных чисел представим в виде $(x - \sqrt{2})(x + \sqrt{2})$; многочлен $x^2 + 1$ неприводим в поле вещественных чисел (и, тем более, в поле рациональных чисел), зато в поле комплексных чисел может быть представлен в виде $(x + i)(x - i)$.

Если некоторый многочлен не является неприводимым в рассматриваемом числовом поле, то он называется *приводимым* в этом поле.

Аналогично простым числам, неприводимые многочлены обладают следующим свойством:

многочлен $p(x)$ над полем T тогда и только тогда неприводим в поле T, когда для любых многочленов $f(x)$ и $g(x)$ над полем T из условия $p(x) \mid f(x) \cdot g(x)$ (т. е. $p(x)$ является делителем для $f(x) \cdot g(x)$) следует, что или $p(x) \mid f(x)$, или $p(x) \mid g(x)$.

Перечисленные результаты дают возможность доказать следующую теорему, аналогичную основной теореме теории чисел:

всякий многочлен над полем T, степень которого не меньше единицы, однозначно разлагается над полем T в произведение неприводимых множителей.

(Доказательство аналогично доказательству основной теоремы теории чисел.)

ЗНАЧЕНИЕ МНОГОЧЛЕНА ПРИ ДАННОМ ЗНАЧЕНИИ НЕИЗВЕСТНОГО; КОРЕНЬ МНОГОЧЛЕНА

До сих пор мы рассматривали операции над многочленами. В этом появляется необходимость при решении (алгебраических) уравнений, что по существу означает отыскание корней многочлена. К этому мы сейчас и перейдем.

Пусть $f(x) = a_0 + a_1x + \ldots + a_nx^n$ — многочлен над полем T, и пусть α — некоторый элемент поля T.

Значением многочлена $f(x)$ при $x = \alpha$ называется число $f(\alpha) = a_0 + a_1\alpha + \ldots + a_n\alpha^n$ (очевидно, что это число принадлежит к полю T).

Число α называется корнем многочлена $f(x)$, если значение $f(x)$ при $x = \alpha$ равно 0.

При вычислении значений многочленов мы часто будем использовать тот факт, что *значение суммы (произведения) двух многочленов при $x = \alpha$ равно сумме (произведению) значений этих многочленов при $x = \alpha$.*

На языке формул это означает, что из соотношений $f(x) + g(x) = h(x)$ и $f(x) \cdot g(x) = k(x)$ следует, что $f(\alpha) + g(\alpha) = h(\alpha)$ и $f(\alpha) \cdot g(\alpha) = k(\alpha)$. Это является следствием того, что действия над неизвестным x производятся точно так же, как над числами из поля T.

Понятие значения многочлена при данном значении неизвестного тесно связано с делением многочленов. Разделим многочлен $f(x)$ на многочлен $(x - \alpha)$. Остаток от деления либо будет иметь степень 0, либо сам окажется равным 0. Таким образом в любом случае остаток есть некоторая постоянная, то есть $f(x) = q(x)(x - \alpha) + r$. Рассмотрим теперь значения обеих частей равенства при $x = \alpha$. Учитывая, что значение (постоянного) многочлена r равно r независимо от α, имеем: $f(\alpha) = q(\alpha)(\alpha - \alpha) + r$. Поскольку второй сомножитель в первом слагаемом есть 0, то и все слагаемое есть 0, откуда $f(\alpha) = r$. Подставляя в результат деления вместо r значение $f(\alpha)$ получаем:

$$\boxed{f(x) = q(x)(x - \alpha) + f(\alpha).}$$

Это так называемая *теорема Безу*.

Из теоремы Безу немедленно следует, что *число α тогда и только тогда является корнем многочлена $f(x)$, когда* $(x - \alpha) \mid f(x)$.

Отсюда, в свою очередь, следует, что *если число α из поля T является корнем многочлена $f(x)$, неприводимого в поле T, то* $\mathit{cm}\,(f) = 1$.

Действительно, если $f(\alpha) = 0$, то $f(x) = q(x)(x - \alpha)$, а значит $\mathit{cm}\,(q) = \mathit{cm}\,(f) - 1 < \mathit{cm}\,(f)$. Поскольку оба сомножителя являются многочленами над полем T, а многочлен $f(x)$ неприводим в поле T, то степени обоих сомножителей не могут быть меньше степени $f(x)$. Как мы видели, степень $q(x)$ меньше степени $f(x)$, следовательно, степень $(x - \alpha)$ не может быть ни меньше, ни, очевидно, больше — а значит, равна — степени $f(x)$. Таким образом $f(x)$ есть многочлен первой степени. Отсюда вытекает также, что неприводимый многочлен имеет максимум один корень (разумеется, в том числовом поле, в котором он неприводим). Согласно данным до сих пор определениям, другие числовые поля не принимаются в рассмотрение, ведь и само значение многочлена берется из исходного поля. Это представляется слишком большим ограничением. Например, многочлен $x^2 - 2$ имеет рациональные коэффициенты и неприводим в поле рациональных чисел (ниже это будет доказано). Согласно сказанному выше, этот многочлен не должен иметь корня, тем не менее, $\sqrt{2}$ является его корнем. Однако утверждение «этот многочлен не должен иметь корня» в данном случае было ошибочным. Мы имеем право сказать, что нет рационального корня (и так оно и есть). В области вещественных чисел этот многочлен имеет корень, но получается, что его даже нельзя подставить в многочлен, так как подставлять можно лишь элементы основного поля T. Однако эта трудность легко устранима, так как коэффициенты многочлена $x^2 - 2$ являются не только рациональными, но одновременно и вещественными числами.

Один корень могут иметь не только неприводимые многочлены. Например, очевидно, что многочлен $(x - 1)^2$ не является неприводимым, и тем не менее имеет только один корень. Правда, этот корень «двойной», так как если представить многочлен в виде $(x-1)(x-1)$, то ясно, что число 1 является корнем каждого из сомножителей. Такие корни называются *кратными корнями* многочлена. Точнее:

число α называется k-кратным корнем многочлена $f(x)$, если $(x-\alpha)^k | f(x)$, но $(x-\alpha)^{k+1} \nmid f(x)$, то есть если существует представление $f(x) = (x-\alpha)^k \cdot g(x)$, причем $g(\alpha) \neq 0$.

Пусть числа $\alpha_1, \alpha_2, \ldots, \alpha_s$ являются корнями многочлена $f(x)$, и их кратности равны, соответственно, $k_1, k_2, \ldots, k_s$. В этом случае говорят, что многочлен $f(x)$ имеет s различных корней, общее же число корней равно $k_1 + k_2 + \ldots + k_s$ (все это, разумеется, относится к некоторому основному полю T).

Докажем теперь, что

общее число корней многочлена $f(x)$ над полем T не превышает его степени.

(Отсюда сразу получается та же верхняя граница и для числа различных корней многочлена.)

Рассмотрим разложение многочлена $f(x)$ над полем T в произведение неприводимых множителей:

$$f(x) = f_1(x) f_2(x) \ldots f_s(x).$$

Предположим, что α является корнем $f(x)$. В результате подстановки получаем:

$$0 = f(\alpha) = f_1(\alpha) f_2(\alpha) \ldots f_s(\alpha).$$

Произведение чисел в правой части может равняться 0 только если хотя бы один из сомножителей равен 0. Таким образом, всякий корень многочлена $f(x)$ является корнем хотя бы одного из многочленов $f_i(x) / i = 1, 2, \ldots, s)$. Мы не принимаем здесь во внимание кратность корней, так как каждый корень появляется в стольких сомножителях, какова его кратность. Выше мы видели, что неприводимый многочлен может иметь не больше одного корня. Отсюда следует, что число корней многочлена $f(x)$ не превышает s. С другой стороны, степень $f(x)$ равна сумме степеней многочленов $f_i(x)$. Поскольку степень каждого из них не меньше единицы, то степень $f(x)$ не меньше s. Таким образом

$$\text{ст}(f) \geq s \geq \text{число корней многочлена } f(x), \text{ что и требовалось доказать.}$$

При рассмотрении кватернионов было отмечено, что для них аналогичное утверждение неверно. Покажем теперь, почему это так.

По правилу умножения многочленов имеем: $(x-i)(x+i) = x^2 + 1$. Подставляя в это равенство $x = j$, слева получаем $(j-i)(j+i) = 1 - k - k + 1 = -2k$, а справа $j^2 + 1 = -1 + 1 = 0$. Таким образом в области кватернионов *значение произведения не обязательно равно произведению значений сомножителей.*

Рассмотрим причину особенного поведения кватернионов. В общем виде произведение $(x+\alpha)(x+\beta)$ может быть вычислено следующим образом:

$$(x+\alpha)(x+\beta)=x^2+(\alpha+\beta)x+\alpha\beta.$$

Подставляя в левую и правую части этого равенства $x=\gamma$, получаем, соответственно:

$$(\gamma+\alpha)(\gamma+\beta)=\gamma^2+\gamma\beta+\alpha\gamma+\alpha\beta; \qquad \gamma^2+\alpha\gamma+\beta\gamma+\alpha\beta.$$

Если оба результата равны, то должно быть $\gamma\beta=\beta\gamma$, что, вообще говоря, не имеет места. (Другими словами, «виной всему» является отсутствие коммутативности). Умножение многочленов было произведено таким образом, как если бы выполнялось равенство $x\beta=\beta x$, то есть мы как бы воспользовались коммутативностью умножения неизвестного на число. Попробуем тогда обойтись без коммутативности; имеем: $(x+\alpha)(x+\beta)=x^2+\alpha x+x\beta+\alpha\beta$. Однако полученное выражение *не является многочленом*. В многочлене неизвестное всегда стоит после «числа». Введение такого умножения привело бы к значительно большим трудностям при вычислениях, чем то обстоятельство, что значение произведения двух многочленов не всегда совпадает с произведением значений сомножителей.

«ОСНОВНАЯ ТЕОРЕМА АЛГЕБРЫ»

Как мы уже видели, число корней многочлена не может превышать его степени. Спрашивается, сколько корней может иметь многочлен *по меньшей мере*. Прежде всего нужно выяснить, о каком числовом поле идет речь. В самом деле, поскольку над любым числовым полем можно указать неприводимый многочлен сколь угодно высокой степени (ниже будет дан такой пример), то многочлен произвольной степени может не иметь в этом поле ни одного корня. С другой стороны, возникает проблема кратности корней. Например, многочлен $(x-1)^n$ имеет только один «различный» корень, то есть в случае любого числового поля может случиться, что многочлен высокой степени имеет лишь один корень. Чтобы избежать этого, условимся, что нас будет интересовать только общее число корней многочлена.

Выбрать основное числовое поле оказывается значительно трудней. Исследования решений уравнений показали, что для уравнений второй, третьей и четвертой степени всегда можно найти, соответственно, 2, 3 или 4 корня (не обязательно различных), если в качестве основного поля выбрать поле комплексных чисел. (К исследованию уравнений мы перейдем позже.) Это явилось причиной, по которой была высказана *«основная теорема алгебры»*, утверждающая, что *всякий многочлен n-ой степени с комплексными коэффициентами имеет в поле комплексных чисел n корней*. Основная теорема эквивалентна утверждению, что всякий многочлен, степень которого не меньше единицы, имеет хотя бы один комплексный корень. (Ниже будет строго сформулировано, в каком смысле эквивалентны эти два утверждения.) Эта теорема действительно является «основной» в классической алгебре, так как если бы она не была справедлива, то нельзя было бы находить корни многочленов, что являлось одной из основных задач классической алгебры (второй основной задачей было «выражение» решений уравнений с помощью радикалов, а третьей — решение систем уравнений). «Основная теорема алгебры» высказывает определенное свойство поля комплексных чисел, и по своему характеру, а также по методу доказательства может быть отнесена к теории функций комплексного переменного. Известно очень много доказательств этой теоремы. Однако, как правило, в этих доказательствах делается упор не на то, что всякий многочлен с комплексными коэффициентами имеет хотя бы один комплексный корень, а доказываются аналогичные утверждения, имеющие лишь другую формулировку. То, что все эти утверждения являются по существу различными формами основной теоремы алгебры, может быть показано уже алгебраическими методами. Теоремы, аналогичные этой, справедливы и для других числовых полей.

Числовое поле K называется алгебраически замкнутым, если всякий многочлен с коэффициентами из K, степень которого не меньше единицы, имеет в поле K хотя бы один корень.

Методами алгебры можно показать, что для любого числового поля K эквивалентны следующие утверждения:

1. *поле K алгебраически замкнуто;*

2. *всякий многочлен над полем K, степень которого не меньше единицы, представим в виде $g(x)(x-\alpha)$, где $g(x)$* и *$(x-\alpha)$ — многочлены над полем K;*

3. *всякий многочлен над полем K, неприводимый в поле K, является многочленом первой степени;*

4. *всякий многочлен над полем K, степень которого не меньме единицы, может быть представлен в виде $a(x-\alpha_1)(x-\alpha_2)\ldots(x-\alpha_n)$, где a, α_1, $\alpha_2, \ldots, \alpha_n$ — числа из K;*

5. *всякий многочлен над полем K, степень которого не меньше единицы, может быть представлен в виде $a\cdot(x-\beta_1)^{k_1}(x-\beta_2)^{k_2}\ldots(x-\beta_s)^{k_s}$, где a принадлежит K, а $\beta_1, \beta_2, \ldots, \beta_s$ — различные числа из K;*

6. *число корней произвольного многочлена над полем K равно его степени;*

Другими словами, если для некоторого числового поля K нам удастся (произвольным способом) доказать одно из перечисленных шести утверждений, то уже методами алгебры можно показать, что в этом случае для рассматриваемого числового поля K справедливы и остальные утверждения. Основная теорема алгебры как раз и утверждает, что поле комплексных чисел обладает перечисленными шестью свойствами.

В некоторых формулировках основной теоремы алгебры высказывается ряд свойств многочленов с вещественными коэффициентами. Эти свойства мы сформулируем следующим образом:

для любого числового поля V эквивалентны следующие утверждения:

2′. *всякий многочлен над полем V, степень которого не меньше единицы, представим в виде $g(x)\,q(x)$, где $g(x)$ и $q(x)$ также являются многочленами над полем V, причем степень $q(x)$ не больше двух;*

3′. *всякий многочлен над полем V, неприводимый в поле V, имеет степень не выше второй;*

4'. всякий многочлен над полем V, степень которого не меньше единицы, может быть разложен над полем V в произведение неприводимых в V сомножителей, степень каждого из которых не больше двух;

5'. всякий многочлен над полем V, степень которого не меньше единицы, может быть представлен в виде произведения целых степеней множителей, каждый из которых является неприводимым в поле V многочленом максимум второй степени;

7'. всякий многочлен нечетной степени над полем V имеет в V хотя бы один корень; для всякого числа α из поля V найдется такое число β из V, что или $\beta^2 = \alpha$ или $\beta^2 = -\alpha$.

(Утверждения, аналогичные приведенным выше, были записаны под теми же номерами.)

В заключение установим связь между числовыми полями V и K:

числовое поле K тогда и только тогда является алгебраически замкнутым, когда найдется такое числовое поле V, которое обладает одним из свойств 2', . . ., 7', и всякий элемент поля K имеет вид $a + bi$, где a и b — числа из V.

МНОГОЧЛЕНЫ С РАЦИОНАЛЬНЫМИ КОЭФФИЦИЕНТАМИ

Пусть

$$f(x) = a_0 + a_1x + \ldots + a_nx^n$$

— многочлен с *рациональными* коэффициентами. Приведем коэффициенты к общему знаменателю и вынесем за скобку наибольший общий делитель их числителей, в результате получим:

$$f(x) = \frac{a}{b}(c_0 + c_1x + \ldots + c_nx^n),$$

где a, b, а также $c_0, c_1, \ldots, c_n$ — целые числа, причем числа c_i взаимно просты. Многочлен, стоящий в скобках, может во многих отношениях (разложение на множители, корни) заменить исходный многочлен. Ввиду этого целесообразно коротко остановиться на изучении многочленов такого типа, получивших название «примитивных многочленов», тем более, что они понадобятся нам впоследствии.

Многочлен $f(x)$ называется примитивным, если

1. его коэффициенты — целые числа;

2. все его коэффициенты в совокупности (но не попарно!) взаимно просты.

Всякий многочлен с рациональными коэффициентами однозначно представим в виде произведения некоторого рационального числа на некоторый примитивный многочлен.

Возможность такого представления мы уже видели; однозначность нетрудно доказывается с помощью теории чисел.

Важным свойством примитивных многочленов является то, что *произведение двух (или большего числа) примитивных многочленов также является примитивным многочленом.*

Ниже мы познакомимся с некоторыми теоремами о многочленах с целыми коэффициентами. Однако достаточно сформулировать эти теоремы для примитивных многочленов, так как это почти не умаляет их общности.

Лемма Гаусса:

если примитивный многочлен разлагается в произведение двух многочленов с рациональными коэффициентами, то он разлагается и в произведение двух примитивных многочленов, которые отличаются от сомножителей первого разложения лишь постоянным множителем, то есть имеют, соответственно, те же степени, что и сомножители первого разложения.

Отсюда сразу следует, что

если примитивный многочлен не разлагается в произведение двух примитивных многочленов меньшей степени, то этот многочлен неприводим в поле рациональных чисел.

Эта лемма имеет большое значение для исследования приводимости многочленов над полем рациональных чисел. В самом деле, как мы видели, всякий многочлен с рациональными коэффициентами можно заменить некоторым примитивным многочленом, после чего следует лишь посмотреть, можно ли разложить этот примитивный многочлен в произведение многочленов с *целыми коэффициентами.*

С помощью *леммы Гаусса* нетрудно сформулировать критерий, с успехом используемый для установления неприводимости примитивных многочленов в поле рациональных чисел.

Критерий Эйзенштейна: если для примитивного многочлена

$$f(x) = a_0 + a_1x + a_2x^2 + \ldots + a_nx^n$$

найдется такое простое число p, что $p \mid a_0$, $p \mid a_1, \ldots, p \mid a_{n-1}$ (тогда очевидно, что $p \nmid a_n$, так как $f(x)$ — примитивный многочлен), но $p^2 \nmid a_0$, то многочлен $f(x)$ неприводим в поле рационалных чисел.

Этот критерий — как видно из самой формулировки — является лишь *достаточным*, но не необходимым условием неприводимости, то есть существуют многочлены, неприводимые в поле рациональных чисел, для которых этот критерий не выполняется.

С помощью критерия Эйзенштейна можно доказать, например, что *для любого натурального числа n найдется многочлен n-ой степени, неприводимый над полем рациональных чисел.*

Таков, например, многочлен $x^n - 2$. Действительно, в этом случае подходит простое число $p = 2$. Поскольку все коэффициенты данного многочлена, кроме старшего, равны 0 или 2, то все они делятся на 2, в то же время постоянный член 2 не делится на $2^2 = 4$. Из неприводимости этого многочлена следует, например, что он не имеет рациональных корней (при $n > 1$), а значит, число $\sqrt[n]{2}$ — иррациональное, и, в частности, иррациональным является и число $\sqrt{2}$ (разумеется, если бы мы хотели доказать *только* это, то можно было бы поступить проще).

РАЦИОНАЛЬНЫЕ КОРНИ МНОГОЧЛЕНОВ; СХЕМА ГОРНЕРА

Для многочленов с целыми коэффициентами часто нелегко бывает установить, имеет ли многочлен целый или рациональный корень. Чтобы не пробовать наугад, целесообразно найти условие, позволяющее установить, какие рациональные числа могут быть корнями данного многочлена.

Если дробь $\frac{u}{v}$, *где* $(u, v) = 1$, *является корнем многочлена*

$$f(x) = a_0 + a_1x + \ldots + a_nx^n$$

с целыми коэффициентами, то $u \mid a_0$ и $v \mid a_n$.

Доказательство:

Предположим, что $\frac{u}{v}$ есть корень многочлена $f(x)$, то есть $f\left(\frac{u}{v}\right) = 0$. После подстановки и умножения на v^n получаем:

$$a_0v^n + a_1v^{n-1}u + \ldots + a_{n-1}vu^{n-1} + a_nu^n = 0.$$

Первые n слагаемых в левой части содержат множитель v, поэтому их сумма делится на v. Поскольку число 0, стоящее в правой части равенства, также делится на v, то и последнее слагаемое в левой части должно делиться на v, то есть $v \mid a_nu^n$. Рассматривая последние n слагаемых в левой части, аналогичным образом получаем, что $u \mid a_0v^n$. Отсюда, если принять во внимание условие $(u, v) = 1$, и следует требуемый результат: $v \mid a_n$, $u \mid a_0$.

Конечно, не всякая такая дробь будет корнем многочлена. Например, единственным корнем многочлена $x-10$ является число 10, хотя указанному выше условию удовлетворяют числа $+1$, -1, $+2$, -2, $+5$, -5, $+10$ и -10.

Обычно в таких случаях приходится подставлять много чисел, поэтому целесообразно производить подстановку таким образом, чтобы выполнять при этом как можно меньше действий. Если сначала выполнить возведение в степень, затем умножение и, наконец, сложение, то в общем случае потребуется произвести $2n-1$ умножений и n сложений. Однако применение теориемы Безу позволяет сократить число необходимых действий (одновременно это дает и другие преимущества). Поскольку речь будет идти о делении, то запишем многочлен по убывающим степеням неизвестного.

Метод Горнера. Пусть

$$f(x) = a_nx^n + a_{n-1}x^{n-1} + \ldots + a_2x^2 + a_1x + a_0$$

и мы хотим определить $f(\alpha)$. По теореме Безу

$$f(x) = g(x)(x - \alpha) + f(\alpha),$$

где, очевидно, $g(x)$ — многочлен $(n-1)$-ой степени:

$$g(x) = b_{n-1}x^{n-1} + \ldots + b_2x^2 + b_1x + b_0.$$

Выполнив умножение в предыдущем равенстве, получаем:

$$f(x) + \alpha g(x) = xg(x) + f(\alpha).$$

Из равенства этих многочленов вытекает равенство их коэффициентов при одинаковых степенях x. Выпишем эти коэффициенты, начиная со старшего:

$$\begin{aligned}
a_n &= b_{n-1}\\
a_{n-1} + \alpha \cdot b_{n-1} &= b_{n-2}\\
a_{n-2} + \alpha \cdot b_{n-2} &= b_{n-3}\\
&\ \ \vdots\\
a_{i+1} + \alpha \cdot b_{i+1} &= b_i\\
&\ \ \vdots\\
a_2 + \alpha \cdot b_2 &= b_1\\
a_1 + \alpha \cdot b_1 &= b_0\\
a_0 + \alpha \cdot b_0 &= f(\alpha).
\end{aligned}$$

Отсюда можно по порядку вычислить все коэффициенты многочлена $g(x)$, причем на последнем шагу мы как раз получим значение $f(\alpha)$. На первом шагу не надо производить никакого действия. На каждом из последующих шагов нужно выполнить одно умножение и одно сложение, в результате и здесь необходимо произвести n сложений, зато умножать нужно n раз вместо $2n - 1$. Преимуществом этого процесса является и то, что по ходу дела мы получаем все коэффициенты частного, поэтому если $f(\alpha) = 0$, то очевидно, что в дальнейшем достаточно разыскивать корни частного.

Другое преимущество этого метода состоит в том, что его легко сделать «механическим», что очень выгодно при работе на вычислительных машинах. Нетрудно заметить, что коэффициенты всякий раз получаются по определенному закону, включая и значение $f(\alpha)$. (Больше того, точно так же можно получить и первый коэффициент, если положить $b_n = 0$). Закон образования коэффициентов можно сформулировать следующим образом: «предыдущий коэффициент нужно умножить на α и прибавить предыдущий коэффициент исходного многочлена». Запишем теперь в одну строку коэффициенты многочлена $f(x)$, а под ними — соответственные коэффициенты частного и значение многочлена $f(\alpha)$. В начале нижней строки запишем также значение α.

	a_n	a_{n-1}	a_{n-2}	....	a_2	a_1	a_0
α	b_{n-1}	b_{n-2}	b_{n-3}	.....	b_1	b_0	$f(\alpha)$

Это так называемая *схема Горнера*. По этой таблице коэффициенты нижнего ряда можно получить следующим образом: первый коэффициент следует перенести из верхней строки; после этого каждый новый коэффициент получается в результате умножения числа, стоящего слева от него, на число α (стоящее слева с краю) и вычитания из полученного результата числа, стоящего сверху. При этом не следует забывать выписывать и *коэффициенты, равные 0!*

Пусть, например, $f(x) = x^3 - 6x^2 + 11x - 6$. Целые корни этого многочлена нужно искать среды чисел $+1, -1, +2, -2, +3, -3, +6$ и -6. Будем подставлять их по очереди:

	1	−6	11	−6
+1	1	−5	6	0

(то есть, число 1 является корнем многочлена, частное есть $x^2 - 5x + 6$). Продолжим процесс уже под вторым рядом:

	1	−5	6
−1	1	−6	12

(то есть −1 не является корнем)

	1	−5	6
2	1	−3	0

(то есть 2 является корнем, а корнем многочлена $x - 3$, полученного в частном, является число 3).

Таким образом, корнями исходного многочлена являются 1, 2 и 3.

Если бы мы не останавливались на каждом шагу, то весь процесс, произведенный механически, выглядел бы следующим образом:

	1	−6	11	−6	
1	1	−5	6	0	1 является корнем
−1	1	−6	12		(эту строку не следует принимать во внимание)
2	1	−3	0		2 является корнем
−2	1	−5			
3	1	0			3 является корнем

МНОГОЧЛЕНЫ ОТ НЕСКОЛЬКИХ НЕИЗВЕСТНЫХ

Рассмотрим теперь не одно, а несколько неизвестных, которые обозначим через $x_1, x_2, \ldots, x_n$. Выражение

$$a \cdot x_1^{i_1} x_2^{i_2} \ldots x_n^{i_n}$$

мы назовем одночленом. Число a из поля T есть коэффициент одночлена, натуральные числа $i_1, i_2, \ldots, i_n$ — показатели степеней соответственных неизвестных.

Показатель, равный единице, не выписывают; если же показатель при каком-либо неизвестном равен 0, то не выписывают и само неизвестное. Так, например, вместо $x^5y^1u^2v^0$ мы будем писать x^5yu^2 (x, y, u, v — неизвестные). Если показатели при всех неизвестных равны 0, то имеем только *константу* из поля T.

Два одночлена называются подобными, если они отличаются только коэффициентами.

Сумма двух подобных одночленов есть одночлен, подобный слагаемым, коэффициент которого равен сумме коэффициентов слагаемых (например, $3x^2uv^3 + 5x^2uv^3 = 8x^2uv^3$). Коэффициент одночлена, равный единице, не выпсиывают (например, $1 \cdot x^2y^3 = x^2y^3$).

Многочленом $f(x_1, x_2, \ldots, x_n)$ от неизвестных $x_1, x_2, \ldots, x_n$ называется сумма конечного числа различных (не подобных) одночленов, составленных из этих неизвестных.

Пример: выражение $x^3yu + 3xu^2 - 5xv^7$ является многочленом от неизвестных x, y, u и v.

Если коэффициенты всех одночленов, присутствующих в многочлене, принадлежат к числовому полю T, то имеем многочлен над полем T (то есть многочлен с коэффициентами из T).

Чтобы сложить два многочлена, достаточно записать сумму всех составляющих их одночленов, а затем привести подобные члены (то есть сложить подобные одночлены).

Пример: $(x^5y - 3uy^2 + 2v^2y) + (xy^2 + 3uy^2 + 3v^2y) = (x^5y + xy^2 + 5v^2y)$.

Чтобы перемножить два одночлена, нужно перемножить их коэффициенты и сложить показатели при одинаковых неизвестных.

Пример: $(2x^2yv)(-3yuv^3) = -6x^2y^2uv^4$.

Чтобы перемножить два многочлена, нужно умножить каждый член (одночлен) первого многочлена на каждый член второго и привести подобные члены.

Можно показать, что так определенные сумма и произведение двух многочленов над полем T также оказываются многочленами над полем T. Кроме того, в системе многочленов над полем T выполнимо вычитание и имеют место свойства действий.

Система $T[x_1, x_2, \ldots, x_n]$ многочленов от неизвестных $x_1, x_2, \ldots, x_n$ с коэффициентами из поля T — относительно введенных операций — называется кольцом многочленов от n неизвестных над полем T.

Под степенью одночлена понимается сумма показателей при всех входящих в его состав неизвестных.

Пример: степень одночлена x^2yuv^3 равна $2+1+1+3=7$.

Под степенью многочлена понимается наибольшая из степеней входящих в его состав одночленов.

Пример: степени одночленов, составляющих многочлен $x^5y + x^2yuv^3 + 3v^6$ равны, соответственно, $5+1=6$, $2+1+1+3=7$ и 6. Наибольшее из этих чисел — число 7. Следовательно, степень многочлена равна 7.

Если все члены многочлена имеют одинаковую степень, то многочлен называется однородным.

Пример: многочлен $x^2y + yuv - v^3$ является однородным многочленом от неизвестных x, y, u, v.

Если многочлен не меняется ни при какой перестановке неизвестных (см. гл. «Комбинаторика», стр. 106), то он называется *симметрическим* многочленом.

Пример: многочлен $x^2+y^2+xy-x^3y^3$ является симметрическим многочленом от неизвестных x и y; в то же время многочлены $x^2+y^2-2xy^2$ и $x+y+x^2y-xy^2$ не являются симметрическими.

Под k-м элементарным симметрическим многочленом от неизвестных $x_1, x_2, \ldots, x_n$ понимается сумма всевозможных произведений, каждое из которых содержит ровно k из этих неизвестных.

k-ый элементарный симметрический многочлен обозначается через σ_k.

Пример: Элементарными симметрическими многочленами от неизвестных x, y, u, v являются

$$\sigma_1 = x+y+u+v; \qquad \sigma_2 = xy+xu+xv+yu+yv+uv;$$

$$\sigma_3 = xyu+xyv+yuv; \qquad \sigma_4 = xyuv.$$

Имеет место так называемая *основная теорема о симметрических многочленах:*

всякий симметрический многочлен из кольца $T[x_1, x_2, \ldots, x_n]$ единственным образом представим в виде многочлена от элементарных симметрических многочленов $\sigma_1, \sigma_2, \ldots, \sigma_n$ с коэффициентами из поля T.

Пример: $x^2+y^2+u^2 = (x+y+u)^2 - 2(xy+xy+yu) = \sigma_1^2 - 2\sigma_1$.

ПРЕОБРАЗОВАНИЕ АЛГЕБРАИЧЕСКИХ ВЫРАЖЕНИЙ

Под целым алгебраическим выражением понимается многочлен от нескольких неизвестных.

Выше были определены правила сложения и умножения многочленов от нескольких неизвестных, тем самым определены и правила сложения и умножения целых алгебраических выражений.

Подстановка. Записывая в целое алгебраическое выражение вместо неизвестных конкретные числа, мы выполняем операцию подстановки.

Рассмотрим, например, многочлен $f(x, y, u, v) = x^5y + x^2yuv^3 + 3v^6$. Произведем подстановку $x = 1, y = 2, u = -5, v = 1$. Полученное значение многочлена обозначим через $f(1, 2, -5, 1)$. Таким образом, $f(1, 2, -5, 1) = 1^5 \cdot 2 + 1^2 \cdot 2 \cdot -5 \cdot 1^3 + 3 \cdot 1^6 = 2 - 10 + 3 = -5$.

Мы будем говорить, что значением алгебраического выражения при данных значениях неизвестных (1, 2, -5, 1) является -5.

Понятие тождества. Если два целых алгебраических выражения равны, то равны и их значения при любых значениях неизвестных. (Верно и обратное утверждение: если значения двух целых алгебраических выражений равны при любых значениях неизвестных, то эти выражения равны между собой). В таких случаях говорят о *тождестве* (или тождественном равенстве). Тождествами являются, например, равенства, выражающие свойства ассоциативности, дистрибутивности и другие. Эти свойства можно сформулировать следующим образом: рассмотрим неизвестные a, b и c: для них выполняются тождества $(a + b) + c = a + (b + c)$ и $(a + b)c = ac + bc$. Если теперь вместо a, b и c подставлять конкретные числовые значения, то мы всякий раз будем получать равенство, называемое числовым тождеством. Как тождество можно рассматривать и формулу Ньютона (см. гл. «Комбинаторика»), так как при ее выводе используются лишь тождества, связанные со сложением и умножением, которые справедливы и для многочленов.

Очень часто используется следующее тождество:

$$\boxed{(a + b) \cdot (a - b) = a^2 - b^2.}$$

В его справедливости нетрудно убедиться, так как $(a + b)(a - b) = a^2 + ab - ba - b^2$.

Аналогичны и следующие тождества:

$$\boxed{\begin{gathered}(a^2 + ab + b^2)(a - b) = a^3 - b^3 \\ (a^3 + a^2b + ab^2 + b^3)(a - b) = a^4 - b^4 \text{ и т. д.}\end{gathered}}$$

Их доказательство можно провести аналогично предыдущему. Поскольку это тождества, то они останутся справедливы и тогда, когда вместо какого-либо из неизвестных подставить целое алгебраическое выражение. В частности, подставляя вместо b выражение $-b$, получаем:

$$(a^2 - ab + b^2)(a + b) = a^3 + b^3$$

$$(a^3 - a^2b + ab^2 - b^3)(a + b) = a^4 + b^4 \text{ и так далее.}$$

Деление. В множестве целых алгебраических выражений деление с остатком, вообще говоря, *не выполнимо*. Может случиться, что при понижении степени одного из неизвестных степень другого повышается. Несмотря на это, всякое целое алгебраическое выражение *единственным образом разлагается в произведение неприводимых (неразложимых) целых алгебраических выражений*. Однако эта теорема имеет сложное доказательство.

Мы видели, что в системе многочленов от одного неизвестного с комплексными коэффициентами всякий неприводимый многочлен имеет степень 1. В данном случае аналогичное утверждение неверно: например, многочлен $a^4 + b^4 + c^4$ неприводим (в произвольном числовом поле). Тем не менее, пользуясь единственностью разложения многоченов от нескольких неизвестных на неприводимые множители, можно говорить о *наибольшем общем делителе целых алгебраических выражений*.

Найдем, например, наибольший общий делитель выражений $a^2 + 2ab + b^2$ и $a^2 - b^2$. Разлагая оба выражения на множители, получаем: $(a + b)(a + b)$ и $(a + b)(a - b)$. Поскольку мы получили многочлены первой степени, то есть заведомо неприводимые, то дальнейшее разложение невозможно. Очевидно, что наибольшим общим делителем является $(a + b)$. Нетрудно видеть также, что наименьшим общим кратным этих двух выражений будет выражение $(a + b)^2(a - b) = a^3 + a^2b - ab^2 - b^3$.

Аналогично тому, как из целых чисел были образованы рациональные, из целых алгебраических выражений можно образовывать рациональные алгебраические выражения или, как их еще называют, рациональные дроби. Таким образом, рациональная дробь представляет собой отношение двух целых алгебраических выражений. Равенство двух рациональных дробей, а также правила операций над ними определяются по тому же принципу, как это было сделано для рациональных чисел.

В дальнейшем для краткости мы будем обозначать целые алгебраические выражения одной буквой, то есть, если f и g — целые алгебраические выражения, то $\frac{f}{g}$ — рациональная дробь. Условимся, однако, считать, что целое выражение g, стоящее в знаменателе, не может быть нулем. Может случитья, что при некоторых значениях неизвестных знаменатель все же обращается в 0. В этом случае результат подстановки не имеет смысла, что не мешает нам, однако, рассматривать данное выражение. Например, можно рассматривать дробное выражение $\frac{x+1}{x-1}$, но нельзя выполнить подстановку $x = 1$, так как тогда выражение теряет смысл.

Дробные выражения $\frac{f}{g}$ и $\frac{u}{v}$ *называются равными* $\left(\text{*обозначается:* } \frac{f}{g} = \frac{u}{v}\right)$, *если* $fv = gu$.

Например, $\dfrac{(x+1)^2}{x^2-1}=\dfrac{x+1}{x-1}$.

Здесь уже оказывается неверным утверждение, что в случае равенства двух выражений равны и их значения при любных значениях неизвестных. Например, при подстановке $x = -1$ дробь в левой части теряет смысл, в то время как дробь в правой части обращается в нуль. Однако нетрудно убедиться, что равенство выполняется при всех (допустимых) значениях неизвестного, при которых обе части равенства имеют смысл.

Сложение и умножение рациональных дробей определяется следующим образом:

$$\frac{f}{g}+\frac{u}{v}=\frac{fv+gu}{gv}; \qquad \frac{f}{g}\cdot\frac{u}{v}=\frac{fu}{gv}.$$

При таком определении справедливы основные свойства операций над рациональными дробями. Кроме того, сложение, вычитание, умножение и деление на выражение, отличное от 0, всегда выполнимы, причем однозначно.

Понятие параметра. Среди алгебраических выражений встречаются такие, в которых некоторые из неизвестных считаются фиксированными. Например, в квадратном трехчлене, то есть в выражении ax^2+bx+c неизвестные a, b и c выполняют иную роль, нежели x. Действительно, в каждом конкретном случае мы подставляем их числовые значения, в зависимости от которых определяем соответствующие значения x.

Неизвестные величины, значения которых задаем мы сами, называются параметрами.

Какие неизвестные следует выбрать в качестве параметров, обычно определяется уже самим подходом к исследованию выражения. Исследование выражений, содержащих параметры, имеет важное значение для решения уравнений. При исследовании выражений, содержащих параметры, многие термины приобретают несколько иной смысл, согласно тому, что данное выражение рассматривается уже как выражение от остальных неизвестных. Например, выражение $7a^2xy$, если ни одно из неизвестных не является параметром, имеет степень 4. Если же рассматривать a как параметр, а x и y считать неизвестными, то имеем выражение второй степени. Если же и x считать параметром, то получим выражение первой степени. Другой пример: выражения $7a^2xy$ и $-3bxy$ не являются подобными, если ни одно из неизвестных не рассматривать как параметр. Если же рассматривать a и b как параметры, то эти выражения следует считать подобными. В этом случае их можно сложить: $7a^2xy - 3bxy = (7a^2-3b)xy$. Операцию внутри скобок уже не удается произвести, так как данные выражения по сути дела все же не являются подобными. Речь идет лишь о том, что с такой точки зрения выражение $(7a^2-3b)$ следует рассматривать как один сомножитель. (Подставляя вместо a и b любые конкретные значения, мы ведь все равно получим одно число.) Выражение $\frac{ax+by}{a+b}$, например, является дробным, если ни одно из неизвестных не рассматривать как параметр. Если же считать a и b параметрами, то это выражение следует рассматривать как целое алгебраическое выражение.

АРИФМЕТИЧЕСКАЯ И ГЕОМЕТРИЧЕСКАЯ ПРОГРЕССИИ

Если каждому натуральному числу n поставить в соответствие некоторое число a_n, то мы получим числовую последовательность, которая обозначается следующим образом:

$$a_1, a_2, \ldots, a_n, \ldots$$

Числа $a_1, a_2, \ldots, a_n, \ldots$ называются *членами последовательности*. Члены последовательности не обязательно различны, последовательностью является, например, и 1, 0, 1, 0, ... Выражение «поставим в соответствие» еще не означает, что мы действительно можем установить некоторое соответствие. Речь идет лишь о том, что такое соответствие *существует*. Числовая последовательность остается последовательностью и в том случае, если вообще нельзя указать закона, по которому образуются ее члены. Пусть, например, a_n равняется температуре 0-го километрового камня в Будапеште в 12 часов дня 1-го января n-го года нашей эры, если эту температуру можно определить, и $a_n = -1000$ в противном случае.

Арифметическая прогрессия. Если последовательность такова, что, вычитая из любого ее члена предыдущий член, мы всякий раз будем получать одно и то же число d, то последовательность называется арифметической прогрессией.

(Это, естественно, не относится к первому члену последовательности).

Таким образом, для арифметической прогрессии характерно то, что $a_{n+1} - a_n = d$, то есть $a_{n+1} = a_n + d$. (Заметим, что недостаточно было бы сказать, что «разность между любыми двумя соседними членами постоянна», так как при вычитании имеет значение, какое число вычитается из другого. При таком определении последовательность 1, 0, 1, 0, ... также была бы арифметической прогрессией.) Если известны первый член и разность d прогрессии, то произвольный ее член можно получить по следующей формуле:

$$\boxed{a_n = a_1 + (n-1)d\,.}$$

Справедливость этого соотношения можно доказать, например, методом полной индукции.

Обозначим через S_n сумму $a_1 + a_2 + \ldots + a_n$. С помощью полной индукции можно доказать также, что

$$\boxed{S_n = n \cdot a_1 + \frac{n(n-1)}{2} \cdot d.}$$

Подставляя вместо $(n-1)d$ разность $a_n - a_1$, получаем:

$$\boxed{S_n = \frac{n}{2}(a_1 + a_n).}$$

Итак, мы установили два соотношения между n, d, a_1, a_n и S_n. Если известны три из этих пяти чисел, то из полученных соотношений можно определить и два другие. Это не вызывает затруднений, если среди известных чисел присутствует n. Если же n неизвестно, то по существу представляются две возможности. Если известны a_1, a_n и d, то n можно вычислить из соотношения $a_n = a_1 + (n-1)d$.

Пример: Пусть первый член прогрессии равен 7, а разность равна 3. Который по счету член прогрессии равен 31, и чему равна сумма членов прогрессии до этого члена? По условию $31 = 7 + (n-1)3$, откуда $n = 9$. Таким образом, 31 является 9-м членом прогрессии. При вычислении n не следует забывать, что *n может быть лишь натуральным числом*. Если бы, например, в нашей задаче было $a_n = 30$, то получилось бы $n = \frac{26}{3}$, что невозможно — такой арифметической прогрессии не существует. Вернемся к нашему примеру: $S_9 = \frac{9}{2} \cdot (7 + 31) = 171$.

Если известны a_1, a_n и S_n, то n можно определить из соотношения $S_n = \frac{n}{2} \cdot (a_1 + a_n)$.

Пусть, например, первый член арифметической прогрессии равен -1, $a_n = 37$ и $S_n = 72$. Чему равны n и d? Имеем: $72 = \frac{n}{2}(-1 + 37)$, откуда $n = 4$. Таким образом $37 = -1 + 3d$, то есть $d = \frac{38}{3}$. Здесь также нужно принять во внимание, что n — натуральное число, в то же время d может быть произвольным числом.

Рассмотрим, наконец, случай, когда известны d, S_n и либо a_1, либо a_n. Оба эти случая рассматриваются, по существу, аналогично. Действительно, рассмотрим первые n членов исходной прогрессии: $a_1, a_2, \ldots, a_{n-1}, a_n$.

Запишем эти числа в обратном порядке: $b_1, b_2, \ldots, b_{n-1}, b_n$, где $b_1 = a_n$, $b_2 = a_{n-1}, \ldots, b_{n-1} = a_2$ и $b_n = a_1$. Полученная последовательность также будет арифметической прогрессией, так как разность $b_{i+1} - b_i = a_{n-i} - a_{n-i+1} = -d$ постоянна. Итак, разность этой прогрессии равна разности исходной прогрессии, взятой с обратным знаком. Очевидно также, что эта прогрессия имеет ту же сумму, так как состоит из тех же самых членов, только записанных в обратном порядке. Таким образом, если известны a_n, d и S_n, то можно рассмотреть арифметическую прогрессию, первый член которой $b_1 = a_n$, разность $d' = -d$, а сумма первых n членов $S'_n = S_n$ (количество членов в обоих случаях одинаково). Если нам удастся отсюда определить количество членов и n-ый член b_n, то тем самым мы получим количество членов исходной прогрессии и, поскольку $a_1 = b_n$, то и ее первый член. Если же известны a_1, d и S_n, то n можно определить из соотношения $S_n = n \cdot a_1 + \frac{n(n-1)}{2} d$. Естественно, что и в этом случае искомая арифметическая прогрессия существует, если n — натуральное число.

Пусть, например, $a_1 = -15$, $d = 3$, а $S_n = -45$. Имеем: $-45 = (-15)n + \frac{n(n-1)}{2} \cdot 3$, откуда получаем, что $n^2 - 11n + 30 = 0$. Решениями этого уравнения второго порядка являются числа 5 и 6. Таким образом, есть две возможности: $a_5 = -15 + 4 \cdot 3 = -3$ или же $a_6 = -15 + 3 \cdot 5 = 0$. Итак, искомая прогрессия имеет вид: $-15, -12, -9, -6, -3, 0, +3, \ldots$, и сумма первых ее пяти или шести членов равна -45.

Известно несколько определений арифметической прогрессии. Рассмотрим, например, три любые соседние члена прогрессии: a_{i-1}, a_i, a_{i+1}. Если разность прогрессии есть d, то $a_{i-1} = a_i - d$ и $a_{i+1} = a_i + d$. Отсюда получаем, что $a_{i-1} + a_{i+1} = 2a_i$, то есть $a_i = \frac{a_{i-1} + a_{i+1}}{2}$. Таким образом, *каждый член арифметической прогрессии (кроме первого) равен среднему арифметическому двух его соседних членов; этим свойством обладают только арифметические прогрессии.* Последнее утверждение также нетрудно доказывается: если $a_i = \frac{a_{i-1} + a_{i+1}}{2}$, то $2a_i = a_{i-1}, + a_{i+1}$, то есть $a_{i+1} - a_i = a_i - a_{i-1}$. Поскольку это выполняется для всех $i > 1$, то $a_n - a_{n-1} = a_{n-1} - a_{n-2} = \ldots = a_3 - a_2 = a_2 - a_1$, и действительно, вычитая из любого члена последовательности предыдущий член, мы всякий раз получаем одну и ту же разность. Выясним теперь, каким образом можно найти n-ый член прогрессии, если известны ее первый член и разность. Воспользуемся соотношением $a_n = a_1 + (n-1)d$. Рассмотрим многочлен $f(x) = dx + a_1$. Имеем: $a_1 = d \cdot 0 + a_1 = f(0)$, $a_2 = d \cdot 1 + a_1 = f(1)$, $\ldots$ $a_n = d(n-1) + a_1 = f(n-1)$, $\ldots$. Таким образом для всякой арифметической прогрессии можно найти такой многочлен первой степени, для которого $a_1 = f(0)$, $a_2 = f(1), \ldots, a_n = f(n-1), \ldots$. Верно и обратное утверждение. Если $f(x)$ — многочлен первой степени, то $f(0), f(1), \ldots, f(n-1), \ldots$ — арифметическая прогрессия. В самом деле, если $f(x) = ax +$

$+ b$, то $f(i+1) - f(i) = a(i+1) + b - (a \cdot i + b) = a$, то есть мы действительно получаем арифметическую прогрессию. Обобщение арифметической прогрессии можно получить, если не предполагать, что $f(x)$ — многочлен первой степени.

Если $f(x)$ — многочлен k-ой степени, то последовательность $f(0)$, $f(1)$, $\ldots$, $f(n-1)$, $\ldots$ называется арифметической прогрессией k-го порядка.

Для арифметической прогрессии k-го порядка также можно найти формулу суммы первых n ее членов. Можно показать, что существует единственный многочлен $g(x)$ $(k+1)$-ой степени, для которого $g(n-1) = f(0) + f(1) + \ldots + f(n-1)$. Если, например, $f(x) = x^2$, то мы получим последовательность квадратов натуральных чисел, то есть, арифметическую прогрессию второго порядка. Ей соответствует многочлен $g(x) = \frac{x(x+1)(2x+1)}{6}$. Таким образом, сумма квадратов первых n натуральных чисел равна $\frac{n(n+1)(2n+1)}{6}$.

Геометрическая прогрессия. Закон, по которому образуются члены геометрической прогрессии, аналогичен закону образования членов арифметической прогрессии.

Если последовательность такова, что в результате деления любого ее члена на предыдущий член мы всякий раз получаем одно и то же число q, то последовательность называется геометрической прогрессией.

Геометрической прогрессией является и последовательность a, 0, $\ldots$, 0, $\ldots$. И хотя здесь только второй член можно делить на предыдущий, тем не менее каждый член прогрессии получается умножением предыдущего члена на 0. Если считать геометрическими прогрессиями и такие последовательности, то справедливо следующее утверждение: *числовая последовательнось тогда и только тогда является геометрической прогрессией, когда всякий член последовательности равен среднему геометрическому двух его соседних членов.* Доказательство проводится точно так же, как и доказательство аналогичного утверждения для арифметических прогрессий; отдельно нужно рассмотреть лишь случай, когда $q = 0$.

Методом полной индукции здесь также доказывается, что

$$\boxed{a_n = a_1 \cdot q^{n-1},}$$

где a_1 — первый, a_n — n-ый член прогрессии, а q — ее знаменатель.

Как найти сумму первых n членов прогрессии? Если $q = 0$, то сумма первых n членов $S_n = a_1$. Если $q = 1$, то очевидно, что $S_n = a_1 + \ldots + a_1 = n \cdot a_1$. Предположим теперь, что $q \neq 0$ и $q \neq 1$. Тогда

$$S_n = a_1 + a_1 q + \ldots + a_1 q^{n-1} = a_1(1 + q + q^2 + \ldots + q^{n-2} + q^{n-1}).$$

Умножим обе части полученного равенства на q:

$$qS_n = a_1(q + q^2 + q^3 + \ldots + q^{n-1} + q^n).$$

Вычитая первое равенство из второго, получаем:

$$(q-1)S_n = qS_n - S_n = a_1(-1+q^n).$$

Поскольку $q \neq 1$, то обе части последнего равенства можно разделить на $(q-1)$:

$$\boxed{S_n = a_1 \cdot \frac{q^n - 1}{q - 1}.}$$

Точно так же, как и в случае арифметической прогрессии, здесь знание трех из пяти чисел — a_1, a_n, n, q и S_n — позволяет с помощью полученных соотношений определить два других. Однако для геометрической прогрессии это оказывается несколько сложнее.

Если среди известных чисел присутствуют q и n, то два недостающих числа всегда удается найти с помощью уравнения первой степени.

Если n известно, а q нет, то существуют две возможности. Если a_1 и a_n известны, то $q = \sqrt[n-1]{\frac{a_n}{a_1}}$, и отсюда можно найти и S_n.

Если известны S_n и a_1, то для q получается уравнение n-ой степени, решение которого вызывает, вообще говоря, серьезные затруднения. В этом случае удается найти лишь приближенное решение. (Знание a_n и S_n приводит, по существу, к такой же задаче.)

Пусть, например, $a_1 = 2$, $a_n = 162$ и $n = 5$. Имеем: $162 = 2q^4$, откуда $q = \sqrt[4]{81} = 3$ (или -3). Следовательно, $S_5 = 2\frac{243-1}{2} = 242$ (или $S_5 = 2\frac{-243-1}{-4} = 122$). Таким образом, мы получаем следующую последовательность: 2, 6, 18, 54, 162 (или 2, -6, 18, -54, 162).

Для иллюстрации второго случая рассмотрим следующий пример: $a_1 = 2$, $S_3 = 62$ (то есть $n = 3$). Имеем: $62 = 2\frac{q^3-1}{q-1}$, отсюда $31 = q^2 + q + 1$, то есть $q^2 + q - 30 = 0$. Это уравнение имеет два корня: 5 и -6. Если $q = 5$, то $a_3 = 2 \cdot 5^2 = 50$. В случае $q = -6$ имеем: $a_3 = 2 \cdot (-6)^2 = 72$. Таким образом мы получаем две возможные последовательности: 2, 10, 50 или 2, -12, 72.

Если n неизвестно, то следует принять во внимание, что n может быть лишь натуральным числом. Встречается однако немало случаев, когда от этого можно отказаться (см. пример ниже). Если q известно, то по существу возможны два случая: или известны a_1 и a_n, или же a_1 и S_n. В обоих случаях n определяется логарифмированием. Рассмотрим пример на упомянутый выше случай.

Через сколько лет удвоится некоторая сумма, вложенная в банк, если в год начисляется 5%? 5%-ное начисление означает, что ежегодно сумма увеличивается в 1,05 раза. По условию $a_n = 2a_1$, таким образом, требуется найти n, для которого $2 = 1{,}05^n$. Отсюда: $n = \frac{\lg 2}{\lg 1{,}05} = \frac{0{,}3010}{0{,}0212} = 14{,}2\ldots$, то есть n не является натуральным числом. Несмотря на это, задача имеет смысл, так как полученный ответ означает, что через 14 лет сумма еще не удвоится, зато через 15 лет уже перейдет этот рубеж.

Казалось бы, что если неизвестны ни n, ни q, то мы имеем самый трудный случай. Однако он довольно легко поддается исследованию. В этом случае должны быть известны a_1, a_n и S_n. Обозначим отношение $\frac{a_n}{a_1}$ через u, а $\frac{S_n}{a_1}$ — через v. Тогда $u = q^{n-1}$ и $v(q-1) = q^n - 1$. Умножая обе части первого равенства на q, получаем: $uq = q^n$. Подставим это выражение для q^n во второе равенство: $v(q-1) = uq - 1$. Это уравнение первой степени относительно q, его решением является $q = \frac{v-1}{v-u}$ (при условии, что $u \neq v$). Отсюда нетрудно определить и n.

Пример: Пусть $a_1 = 3$, $a_n = 96$ и $S_n = 189$. Тогда $u = \frac{96}{3} = 32$, $v = \frac{189}{3} = 63$. Отсюда $q = \frac{63-1}{63-32} = \frac{62}{31} = 2$. Поскольку $u = q^{n-1}$, то $32 = 2^{n-1}$, откуда $n - 1 = 5$, то есть $n = 6$. Таким образом, искомая геометрическая прогрессия имеет следующий вид: 3, 6, 12, 24, 48, 96.

Заметим, что если $u = v$, то $v = 1$, то есть $a_1 = a_n = S_n$ (в случае $a_1 = 0$ q и n могут быть произвольными). Если $a_1 \neq 1$, то должно быть $q^{n-1} = 1$ и $q^n - 1 = q - 1$. Последнее равенство вытекает из предыдущего. Таким образом, подходят все значения q и n, для которых $q^{n-1} = 1$. Если q может принимать только вещественные значения, то $q = 1$ или $q = -1$. В первом случае из $S_n = na_1$ следует, что $n = 1$. Если же $q = -1$, то n может быть любым нечетным числом.

УРАВНЕНИЯ

УРАВНЕНИЯ ПЕРВОЙ И ВТОРОЙ СТЕПЕНЕЙ

Рассмотрим многочлен $f(x)$ с коэффициентами из некоторого числового поля T.

Запись $f(x) = 0$ называют алгебраическим уравнением. Если $cm(f) = n$, то имеем уравнение n-ой степени.

Решение алгебраического уравнения означает разыскание корней многочлена, стоящего в левой части, то есть разыскание чисел α, для которых $f(\alpha) = 0$. Использованное выше (привычное) обозначение несколько искажает смысл уравнения. Дело в том, что при рассмотрении многочленов равенство $f(x) = 0$ означало, что многочлен $f(x)$ равен нулевому многочлену, то есть все его коэффициенты равны 0. Во избежание недоразумений при исследовании уравнений мы будем обозначать равенство многочленов тремя чертами: $f(x) \equiv g(x)$.

Общий вид уравнения первой степени: $ax + b = 0$.

1. Если $a = b = 0$, то решением уравнения является любое число.
2. Если $a = 0$, $b \neq 0$, то уравнение не имеет решений.
3. Если $a \neq 0$, то решение уравнения имеет вид $-\dfrac{b}{a}$.

Общий вид уравнения второй степени: $ax^2 + bx + c = 0$ (где $a \neq 0$).

Разложение квадратного трехчлена на множители. Известный способ решения квадратного уравнения состоит в дополнении квадратного трехчлена до полного квадрата. Если же мы хотим разложить квадратный трехчлен на множители или установить связь между его корнями и коэффициентами, то можно избрать и более короткий путь. Разделим трехчлен $f(x) = ax^2 + bx + c$ на $a(x - \alpha)$, где α — произвольное число. Очевидно, что частным будет многочлен вида $(x - \beta)$, а в остатке получится $f(\alpha)$:

4. $f(x) \equiv a(x - \alpha)(x - \beta) + f(\alpha)$.

Подставляя в это тождество $x = \beta$, получаем: $f(\beta) = f(\alpha)$. Следовательно, если α является корнем трехчлена, то и β будет корнем. Таким образом, если уравнение второй степени имеет корень в поле T, то оно имеет и второй корень. В этом случае справедливо следующее соотношение:

5. $\boxed{f(x) \equiv a(x - \alpha)(x - \beta).}$

Это позволяет записать исходное уравнение в виде

$$a(x - \alpha)(x - \beta) = 0,$$

где трехчлен в левой части разложен на множители. Из соотношения 5 после выполнения умножения получаем:

6. $ax^2 + bx + c \equiv ax^2 - a(\alpha + \beta)x + a\alpha\beta$.

Коэффициенты при одинаковых степенях неизвестного равны, поэтому для $(\alpha + \beta)$ и $\alpha\beta$ получаем уравнения первой степени, откуда

7. $\boxed{\alpha + \beta = -\dfrac{b}{a} \qquad \alpha\beta = \dfrac{c}{a}.}$

Эти формулы выражают связь между коэффициентами и корнями квадратного уравнения. Поскольку разность квадратов двух чисел равна произведению их разности на сумму, то

$$(\alpha+\beta)^2-(\alpha-\beta)^2=(2\alpha)(2\beta)=4\alpha\beta\,.$$

Отсюда:

$$(\alpha-\beta)^2=(\alpha+\beta)^2-4\alpha\beta=\left(-\frac{b}{a}\right)^2-4\frac{c}{a}=\frac{b^2-4ac}{a^2}\,.$$

Извлекая из обеих частей этого равенства квадратный корень, получаем:

8. $\alpha-\beta=\dfrac{\sqrt{b^2-4ac}}{a}\,.$

На самом деле для квадратного корня получается два значения: $(\alpha-\beta)$ и $(\beta-\alpha)$. Если же условиться, какое из значений рассматривается (например, в вещественном случае — положительное), то, поменяв в случае необходимости местами α и β, мы действительно получим в левой части $\alpha-\beta$. Из первого равенства 7 и из равенства 8 после сложения (вычитания) и деления на 2 получаем:

$$\boxed{\alpha=\frac{-b+\sqrt{b^2-4ac}}{2a}\,;\qquad \beta=\frac{-b-\sqrt{b^2-4ac}}{2a}\,.}$$

Таким образом, если уравнение второй степени имеет решения, то они даются этими формулами. Простой подстановкой можно убедиться, что это действительно решения уравнения.

Дискриминант. Выражение $D=b^2-4ac$, стоящее под радикалом, называется дискриминантом уравнения второй степени.

Если коэффициенты уравнения — вещественные числа, то характер решений определяется дискриминантом:

1. если $D>0$, то уравнение имеет два различных решения;
2. если $D=0$, то уравнение имеет одно решение кратности 2;
3. если $D<0$, то уравнение не имеет вещественных решений (однако в поле комплексных чисел оно имеет два решения, сопряженных друг другу).

Если коэффициенты — комплексные числа, то это уравнение всегда имеет два решения, так как из комплексного числа всегда можно извлечь квадратный корень.

Пример. Найдем решения уравнения $10x^2-29x+10=0$. Сначала вычислим корень из дискриминанта:

$$\sqrt{29^2-4\cdot10\cdot10}=\sqrt{841-400}=\sqrt{441}=21\,.$$

Отсюда находим решения:

$$\alpha = \frac{29+21}{20} = \frac{5}{2}\,; \qquad \beta = \frac{29-21}{20} = \frac{2}{5}\,.$$

Для уравнения $x^2 - 2x + 2 = 0$ имеем: корень из дискриминанта есть $\sqrt{4-8} = \sqrt{-4} = 2i$, следовательно, решениями являются

$$\frac{2+2i}{2} = 1+i \qquad \text{и} \qquad \frac{2-2i}{2} = 1-i\,.$$

УРАВНЕНИЕ ТРЕТЬЕЙ СТЕПЕНИ

Общий вид уравнения третьей степени:

1. $ax^3 + bx^2 + cx + d = 0$ где $(a \neq 0)$

ФОРМУЛА КАРДАНО

Если выполнить подстановку $x = \left(y - \dfrac{b}{3a}\right)$, то второй член в уравнении исчезнет:

$$\begin{aligned}
ax^3 &= a\left(y - \frac{b}{3a}\right)^3 = ay^3 - by^2 + \frac{b^2}{3a}\,y - \frac{b^3}{27a^2}\\
bx^2 &= b\left(y - \frac{b}{3a}\right)^2 = \qquad by^2 - \frac{2b^2}{3a}\,y + \frac{b^3}{9a^2}\\
cx &= c\left(y - \frac{b}{3a}\right) = \qquad\qquad cy - \frac{bc}{3a}\\
d &= \qquad\qquad\qquad\qquad d.
\end{aligned}$$

Складывая эти равенства, получаем:

$$0 = ax^3 + bx^2 + cx + d = \qquad ay^3 + \left(c - \frac{b^2}{3a}\right)y + \left(d - \frac{bc}{3a} + \frac{2b^3}{27a^2}\right).$$

После деления на a приходим к следующему уравнению:

$$y^3 + py + q = 0,$$

где p и q могут быть выражены через a, b, c и d с помощью четырех арифметических действий, то есть p и q принадлежат к тому же числовому полю, из которого взяты коэффициенты исходного уравнения (или многочлена). Если нам удастся найти решения полученного уравнения, то из них можно будет определить и решения исходного уравнения. Решения полученного уравнения третьей степени будем искать в виде $y = u + v$. Подставим это выражение в уравнение:

$$(u+v)^3 + p(u+v) + q \equiv u^3 + v^3 + 3uv(u+v) + p(u+v) + q \equiv$$
$$\equiv (u^3 + v^3 + q) + (3uv + p)(u+v)\,.$$

Если удастся найти такие u и v, что $u^3 + v^3 + q = 3uv + p = 0$, то $y = u + v$ будет решением рассматриваемого уравнения. Итак, для u и v мы имеем два соотношения:

$$2.\ u^3 + v^3 = -q\,; \qquad 3.\ uv = -\frac{p}{3}\,.$$

Возводя равенство 3 в куб, получаем: $u^3v^3 = -\left(\frac{p}{3}\right)^3$. Используя связь между коэффициентами и корнями квадратного трехчлена, приходим к выводу, что u^3 и v^3 являются решениями квадратного уравнения $t^2 + qt - \left(\frac{p}{3}\right)^3 = 0$, откуда:

$$u^3 = -\frac{q}{2} + \sqrt{\left(-\frac{q}{2}\right)^2 + \left(\frac{p}{3}\right)^3}\,; \qquad v^3 = \frac{q}{2} - \sqrt{\left(-\frac{q}{2}\right)^2 + \left(\frac{p}{3}\right)^3}\,.$$

Отсюда как для u, так и для v получаем три значения, итого 9 пар значений. Очевидно, что каждая из 9-ти пар удовлетворяет равенству 2. Однако мы видели, что $y = u + v$ является искомым решением, если для u и v выполняется и равенство 3. Обозначим полученные значения через u_1, u_2, u_3 и $v_1, v_2\, v_3$. Если в качестве основного поля рассмотреть поле комплексных чисел, то, производя извлечение кубического корня в тригонометрической форме (и перенумеровав в случае необходимости значения корня), получаем: $u_2 = \varrho u_1$, $u_3 = \varrho^2 u_1$; $v_2 = \varrho^2 v_1$, $v_3 = \varrho v_1$, где $\varrho = \cos 120° + i \cdot \sin 120° = -\frac{1}{2} + i\,\frac{\sqrt{3}}{2}$ $\left(\text{и } \varrho^2 = -\frac{1}{2} - i \cdot \frac{\sqrt{3}}{2}\right)$. Итак, $(u_1v_1)^3 = -\left(\frac{p}{3}\right)^3$, откуда получаем три значения для u_1v_1: $-\frac{p}{3}$, $\varrho\left(-\frac{p}{3}\right)$ и $\varrho^2\left(-\frac{p}{3}\right)$. В соответствии с этим равенству 3 будет удовлетворять или u_1v_1, или u_1v_2, или же u_1v_3. Следовательно из 9 пар значений можно выбрать одну, произведение чисел которой равно $-\frac{p}{3}$. Предположим, что это как раз пара u_1, v_1. Тогда для u_2 и для u_3 найдутся единственные значения v — ими будут как раз v_2 и v_3 — такие, что соответствующие произведения также будут равны $-\frac{p}{3}$.

Таким образом мы нашли три решения уравнения третьей степени:

$$y_1 = u_1 + v_1; \qquad y_2 = u_2 + v_2; \qquad y_3 = u_3 + v_3.$$

Выражая их через коэффициенты, получаем:

$$\boxed{y = \varrho\sqrt[3]{-\frac{q}{2} + \sqrt{\left(-\frac{q}{2}\right)^2 + \left(\frac{p}{3}\right)^3}} + \varrho^2\sqrt[3]{-\frac{q}{2} - \sqrt{\left(-\frac{q}{2}\right)^2 + \left(\frac{p}{3}\right)^3}},}$$

где вместо ϱ нужно подставить, соотвественно, значения $1, -\frac{1}{2} + i \cdot \frac{3}{2}$ и $-\frac{1}{2} - i \cdot \frac{3}{2}$. Это так называемая *формула Кардано*, дающая решения уравнения третьей степени.

КАЗУС НЕПРИВОДИМОСТИ

Решением уравнений третьей степени стали заниматься, когда еще не были известны комплексные числа. Рассмотрим случай, когда коэффициенты исходного уравнения — вещественные числа. Выше было показано, что коэффициенты p и q полученного уравнения принадлежат к тому же числовому полю, что и коэффициенты исходного уравнения, поэтому в данном случае они являются вещественными числами. Как и в случае квадратного уравнения, число под радикалом называется здесь *дискриминантом*: $D = \left(-\frac{q}{2}\right)^2 + \left(\frac{p}{3}\right)^3$. «Характер» решений определяется согласно тому, является ли D положительным, равным 0, или отрицательным.

1. Если $D > 0$, то u^3 и v^3 — вещественные числа, но тогда кубические корни из них имеют по одному вещественному значению. Поскольку и произведение uv есть вещественное число, то можно предположить, что вещественными значениями кубических корней являются, например, u_1 и v_1. Отсюда: $y_1 = u_1 + v_1$;

$$y_2 = \varrho u_1 + \varrho^2 v_1 = -\frac{1}{2}(u_1 + v_1) + i \cdot \frac{\sqrt{3}}{2}(u_1 - v_1);$$

$$y_3 = \varrho^2 u_1 + \varrho v_1 = -\frac{1}{2}(u_1 + v_1) - i \cdot \frac{\sqrt{3}}{2}(u_1 - v_1).$$

Поскольку $u^3 - v^3 = 2\sqrt{D} \neq 0$ (так как $D > 0$), то $u^3 \neq v^3$, а значит $u_1 \neq v_1$. Следовательно, ни y_2, ни y_3 не являются вещественными числами. Таким образом, в этом случае *уравнение третьей степени имеет один вещественный и два сопряженных комплексных корня.*

2. В случае $D = 0$ получается результат, аналогичный предыдущему, с той разницей, что здесь $u_1 = v_1$, поэтому корнями уравнения являются $y_1 = u_1 + v_1$, $y_2 = y_3 = -\frac{1}{2}(u_1 + v_1)$. Таким образом, в этом случае *уравнение третьей степени имеет три вещественных корня, причем (по крайней мере) два из них равны между собой.* (Может случиться, что все три корня равны.)

3. Если $D < 0$, то очевидно, что u^3 и v^3 — сопряженные комплексные числа. Для извлечения кубических корней запишем их в тригонометрической форме. Квадрат модуля комплексного числа $\alpha + \beta i$ равен $\alpha^2 + \beta^2$. Поскольку $D < 0$, то $u^3 = -\frac{q}{2} + i \cdot \sqrt{-D}$ и $v^3 = -\frac{q}{2} - i \cdot \sqrt{-D}$, откуда квадрат их модуля

$$\left(-\frac{q}{2}\right)^2 - D = \left(-\frac{q}{2}\right)^2 - \left(\frac{q}{2}\right)^2 - \left(\frac{p}{3}\right)^3 = -\left(\frac{p}{3}\right)^3.$$

(Это число только выглядит отрицательным; на самом же деле очевидно, что D может быть отрицательным только если и p — отрицательное число). Отсюда получаем:

$$|u| = |v| = \sqrt{-\frac{p}{3}}.$$

Поскольку $uv = -\frac{p}{3}$ — вещественное число, то аргументы чисел u и v отличаются лишь знаком. С другой стороны, их модули равны, следовательно, эти числа сопряжены друг другу: $v_1 = \overline{u}_1$. Если использовать также, что $\varrho^2 = \overline{\varrho}$, то получим:

$$v_2 = \varrho^2 v_1 = \overline{\varrho u_1} = \overline{u}_2; \qquad v_3 = \varrho v_1 = \overline{\varrho^2 u_1} = \overline{u}_3.$$

Таким образом корни уравнения имеют следующий вид:

$$y_1 = u_1 + \overline{u}_1; \qquad y_2 = u_2 + \overline{u}_2; \qquad y_3 = u_3 + \overline{u}_3,$$

причем все три корня — *вещественные.*

Покажем, что эти корни различны. Предположим, например, что $y_2 = y_3$, тогда

$$0 = y_2 - y_3 = \varrho u_1 + \varrho^2 \overline{u}_1 - \varrho^2 u_1 - \varrho \overline{u}_1 = \varrho(u_1 - \overline{u}_1) - \varrho^2(u_1 - \overline{u}_1) =$$
$$= (\varrho - \varrho^2)(u_1 - \overline{u}_1).$$

Произведение может равняться 0 только если хотя бы один из сомножителей равен 0. Однако очевидно, что $\varrho \neq \varrho^2$. Не может иметь места и равенство $u_1 = \overline{u}_1$, так как в этом случае u_1, значит и u_1^3 были бы вещественными числами, что возможно лишь при $D = 0$ (а мы предположили, что $D < 0$).

Аналогично можно показать, что все три корня попарно различны. Таким образом, мы пришли к довольно неожиданному результату: как раз в том случае, когда уравнение третьей степени имеет три различных вещественных корня, разыскать эти корни можно лишь с помощью комплексных чисел. В случае квадратного уравнения отрицательность дискриминанта не вызывала затруднений, так как там просто можно было сказать, что уравнение не имеет корней. Здесь же корни существуют, только их не удается найти без использования комплексных чисел. Этот случай, получивший название *«казуса неприводимости»*, и являлся одним из тех, что привел к необходимости введения комплексных чисел.

Вернемся теперь к извлечению кубического корня из u^3 и v^3, представленных в тригонометрической форме. Соответствующие модули мы уже нашли, обратимся теперь к аргументам. Обозначим $\arg u^3$ через φ, тогда $\cos\varphi = -\dfrac{q}{2\left(\sqrt{-\dfrac{p}{3}}\right)^3}$ Отсюда следует, что аргументы чисел u_1, u_2 и u_3 равны, соотвественно, $\dfrac{\varphi}{3}$, $\dfrac{\varphi+2\pi}{3}$ и $\dfrac{\varphi+4\pi}{3}$. Таким образом корни уравнения принимают следующий вид:

$$y_1 = 2\sqrt{-\frac{p}{3}}\cos\frac{\varphi}{3},$$

$$y_2 = 2\sqrt{-\frac{p}{3}}\cos\frac{\varphi+2\pi}{3},$$

$$y_3 = 2\sqrt{-\frac{p}{3}}\cos\frac{\varphi+4\pi}{3},$$

где $\cos\varphi = -\dfrac{q}{2\left(\sqrt{-\dfrac{p}{3}}\right)^3}$ (Следует, разумеется, учесть, в который из квадрантов попадает φ, однако это приводит разве что к необходимости поменять местами индексы при y_2 и y_3).

Рассмотрим, например, уравнение $y^3 - 21y - 20 = 0$. В данном случае $p = -21$, $q = -20$, следовательно, $-\dfrac{p}{3} = 7$, $-\dfrac{q}{10} = 10$. Отсюда: $\cos\varphi = \dfrac{10}{\sqrt{7^3}}$, то есть $\varphi \sim 53°42'$. Этому соответствуют три угла: $17°54'$, $137°6'$ и $257°54'$. Одсюда находим значения корней: 5,04; —3,93 и —1,11. (Эти значения, естественно, приближенные; корнями уравнения являются 5, —4 и —1.)

УРАВНЕНИЕ ЧЕТВЕРТОЙ СТЕПЕНИ

Общий вид уравнения четвертой степени: $ax^4 + bx^3 + cx^2 + dx + c = 0$ $(a \neq 0)$, где коэффициенты принадлежат к некоторому числовому полю. Это уравнение также можно привести к более простому виду. Если, аналогично тому, как это было сделано для уравнения третьей степени, разделить данное уравнение на a и произвести подстановку $y = x - \frac{b}{4a}$, то мы придем к уравнению следующего вида:

$$y^4 + py^2 + qy + r = 0.$$

Коэффициенты этого уравнения принадлежат к тому же числовому полю, что и коэффициенты исходного уравнения. Решение уравнения четвертой степени уже нецелесообразно записывать в виде формулы, поэтому обычно рассматривают лишь процесс, приводящий к этому решению. Попытаемся представить многочлен, стоящий в левой части, в виде разности квадратов двух многочленов:

$$y^4 = py^2 + qy + r \equiv (y^2 + u)^2 - (Ay + B)^2 \equiv$$
$$\equiv y^4 + (2u - A^2)y^2 + (-2AB)y + (u^2 - B^2).$$

Равенство двух многочленов означает равенство их коэффициентов при одинаковых степенях неизвестного, то есть

$$p = 2u - A^2, \qquad q = -2AB, \qquad r = u^2 - B^2.$$

Отсюда после перестановки членов получаем: $A^2 = 2u - p$, $2AB = -q$, $B^2 = u^2 - r$. Определяя из этих соотношений двумя способами $4A^2B^2$, приходим к следующему равенству:

$$4(2u - p)(u^2 - r) = (-q)^2.$$

После элементарных преобразований получаем:

$$8u^3 - 4pu^2 - 8ru + (4pr - q^2) = 0.$$

Это уравнение третьей степени относительно u, получившее название *кубической резольвенты* исходного уравнения четвертой степени. Такое уравнение третьей степени мы уже умеем решать. Если же известен один из его корней u, то можно вычислить и значения A и B: $A = \sqrt{2u - p}$, $B = \sqrt{u^2 - r}$. Для каждого из радикалов, естественно, получаются два значения, однако, если принять во внимание условие $2AB = -q$, то два значения следует отбросить (на самом деле подходит любое из них; в зависимости от того, которые из них мы выберем, изменится лишь порядок сомножителей в записанном ниже разложении). Зная u, A и B, многочлен, стоящий в левой части уравнения четвертой степени, можно разложить на множители следующим образом:

$$(y^2 + u)^2 - (Ay + B)^2 \equiv (y^2 - Ay + u - B)(y^2 + Ay + u + B).$$

Поскольку произведение может равняться 0 только если хотя бы один из сомножителей равен 0, то корни уравнения четвертой степени совпадают с корнями следующих двух квадратных уравнений:

$$y^2 - Ay + u - B = 0 \qquad \text{и} \qquad y^2 + Ay + u + B = 0.$$

Заметим, что если коэффициенты исходного уравнения — вещественные числа, то и кубическая резольвента будет иметь хотя бы один вещественный корень, который может быть использован для дальнейших вычислений.

Рассмотрим, например, уравнение $y^4 - 42y^2 - 64y + 105 = 0$. Имеем: $-42 = 2u - A^2$, $-64 = -2AB$ и $105 = u^2 - B^2$, то есть, $A^2 = 2(u + 21)$, $AB = 32$ и $B^2 = u^2 - 105$. Отсюда: $2(u + 21)(u^2 - 105) = 32^2$, то есть

$$u^3 = 21u^2 - 105u - 2717 = 0.$$

Подставляя $u = v - 1$, получаем

$$v^3 - 252v - 1296 = 0.$$

Вычислим дискриминант этого уравнения:

$$D = \left(\frac{q}{2}\right)^2 + \left(\frac{p}{3}\right)^3 = 648^2 + (-84)^3 = 419904 - 592704 = -172800 =$$
$$= 57600(-3) = (240)^2 - 3.$$

Итак, одним из корней этого уравнения третьей степени является $\sqrt[3]{648 + 240\sqrt{-3}} + \sqrt[3]{648 - 240\sqrt{-3}}$. Такие кубические корни в некоторых случаях удается вычислить. Мы имеем как раз один из таких случаев. Приведем лишь результат:

$$(9 - \sqrt{-3})^3 = 729 - 243\sqrt{-3} - 81 + 3\sqrt{-3} = 648 + 240\sqrt{-3}.$$

Аналогично, $(9 + \sqrt{-3})^3 = 648 - 240\sqrt{-3}$. Поскольку $(9 + \sqrt{-3})(9 - \sqrt{-3})$ — вещественное число, то эти значения кубических корней нам подходят. Следовательно, одним из корней уравнения третьей степени является $(9 + \sqrt{-3})(9 - \sqrt{-3}) = 18$. Соответствующий корень кубической резольвенты $u = v - 7 = 18 - 7 = 11$. В этом случае $A^2 = 2 \cdot (11 + 21) = 64$, $B^2 = 121 - 105 = 16$. Если учесть, что произведение AB положительно, то получаем, что A и B должны иметь одинаковые знаки. Пусть оба они будут положительны : $A = 8$, $B = 4$. Отсюда для исходного уравнения четвертой степени получаем :

$$(y^2 - 8y + 15)(y^2 + 8y + 7) = 0.$$

Корнями первого сомножителя являются 5 и 3, корни второго равны -1 и -7. Таким образом, уравнение четвертой степени имеет следующие корни : -7, -1, 3 и 5.

ДРУГИЕ ТИПЫ УРАВНЕНИЙ

Краткий обзор. В то время, как решение квадратного уравнения было известно еще древним грекам, решения уравнений третьей и четвертой степеней стали известными лишь в XVI веке. Решение уравнения третьей степени связывают с именем Н. Тарталья однако, общая формула решения была впервые дана Дж. Кардано в его книге, появившейся в 1539 году. Метод решения уравнений четвертой степени был впервые указан Л. Феррари. Немало других математиков пыталось найти формулы решения уравнений более высокой степени, пока, наконец, отчасти Руффини, а затем норвежский математик Абель не показали, что уравнение пятой степени уже неразрешимо аналогичным образом. (При рассмотрении теории поля мы точнее сформулируем, что понимается под решением уравнения.)

Дробные алгебраические уравнения. К целым алгебраическим уравнениям приводятся и так называемые дробные алгебраические уравнения, в которых неизвестное может присутствовать и в знаменателе. Такие уравнения после умножения обеих частей на наименьшее общее кратное знаменателей превращаются в целое алгебраическое уравнение. Следует, однако, учитывать, что поскольку в исходном уравнении неизвестное может присутствовать и в знаменателе, то может случиться, что при подстановке корней полученного уравнения в исходное уравнение знаменатель обращается в 0. Такой корень не является корнем исходного уравнения. По этой причине простой подстановкой следует всегда проверять, является ли полученное число корнем исходного уравнения.

Иррациональные уравнения. В заключение остановимся на так называемых иррациональных уравнениях. Так называются уравнения, в которых неизвестное присутствует под знаком радикала целой степени.

При решении таких уравнений возникают две проблемы. Во-первых, необходимо следить за тем, чтобы решение не вывело нас за пределы основного числового поля (в вещественном случае, например, корень квадратный из отрицательного числа). Вторая, значительно более существенная проблема является следствием неоднозначности извлечения корня.

Решим, например, уравнение $\sqrt{4x+5} + x = 0$. Если для некоторого x это равенство выполняется, то должны выполняться также равенства $\sqrt{4x+5} = -x$ и $4x + 5 = x^2$. Таким образом мы свели решение исходного уравнения к решению уравнения $x^2 - 4x - 5 = 0$. Его корнями являются число 5 и -1. Подставляя эти значения неизвестного в исходное уравнение, получаем, соответственно, $\sqrt{25}+5$ и $\sqrt{1}-1$. Однако квадратный корень может иметь два значения. Для однозначности целесообразно условиться, что в вещественном случае корень кдадратный всегда обозначает положительное значение. В этом случае $\sqrt{25}=5$ и $\sqrt{1}=1$, и, следовательно, 5 не является корнем исходного уравнения, а -1 корень. (Нетрудно убедиться, что число 5 является корнем уравнения $-\sqrt{4x+5}+x=0$).

В вещественном случае нетрудно определить, какое из значений квадратного корня следует взять. Однако в общем случае приходится извлекать корень из комплексных чисел. Обозначим, например, одно из значений корня квадратного из комплексного числа $a+bi$ через $c+di$, тогда «другим» значением будет $-c-di$. Необходимо указать прием, при помощи которого можно было бы раз и навсегда выбрать то из двух значений, которое обозначается через $\sqrt{a+bi}$ (тогда другое значение будет обозначаться через $-\sqrt{a+bi}$). Этот прием может быть произвольным, однако, по аналогии с вещественным случаем, принято следующее определение:

$$\sqrt{a+bi}=c+di, \quad \text{при} \quad c>0$$
$$\text{или, если } c=0, \quad \text{и} \quad d>0.$$

Из двух чисел, $c+di$ и $-c-di$, указанным свойством обладает только одно. В вещественном случае это определение совпадает с данным выше. Следует, однако, проявлять осторожность, так как использование соотношений, справедливых для вещественных чисел, здесь может привести к ошибочным выводам. Если, например, a и b — положительные вещественные числа, то $\sqrt{ab}=\sqrt{a}\cdot\sqrt{b}$. Действительно, возводя обе части в квадрат, получаем ab, поэтому различие может быть лишь в знаке, но оба числа положительны, следовательно, равенство действительно выполняется. Рассмотрим теперь следующие преобразования: $1=\sqrt{1}=\sqrt{(-1)(-1)}= =\sqrt{-1}\cdot\sqrt{-1}=i\cdot i=-1$. Поскольку в результате мы получили равенство $1=-1$, что невозможно, то очевидно, что где-то была допущена ошибка. Второе и последнее равенства, очевидно, справедливы. В первом и предпоследнем равенстве мы воспользовались определением квадратного корня, поэтому ошибочным может быть только равенство $\sqrt{(-1)(-1)}= =\sqrt{-1}\cdot\sqrt{-1}$. Действительно, из равенств $\sqrt{(-1)(-1)}=\sqrt{1}=1$ и $\sqrt{-1}\cdot\sqrt{-1}=i\cdot i=-1$ получаем, что $\sqrt{(-1)(-1)}=-\sqrt{-1}\cdot\sqrt{-1}$. В самом деле, для любых двух комплексных чисел z и w корень $\sqrt{z\cdot w}$ равняется или $(\sqrt{z}\cdot\sqrt{w})$ или же $(-\sqrt{z}\cdot\sqrt{w})$. Когда показатель корня есть натуральное число, большее двух, то чтобы определить, какое комплексное число обозначает знак радикала, целесообразно найти аргумент подкоренного числа. Нетрудно убедиться, что в случае, когда показатель корня равен 2, приводимое ниже определение совпадает с определением квадратного корня.

Для любого комплексного числа z через $\sqrt[n]{z}$ обозначается то комплексное число, для которого $-\frac{\pi}{2} < \arg(\sqrt[n]{z}) \leq \frac{\pi}{2}$.

Встречаются, конечно, и уравнения других типов. К их классификации мы вернемся при изучении теории поля.

УРАВНЕНИЯ, СОДЕРЖАЩИЕ ПАРАМЕТРЫ; ПРЕОБРАЗОВАНИЕ УРАВНЕНИЙ

Уравнения, содержащие параметры. Общий вид квадратного уравнения: $ax^2 + bx + c = 0$. При его решении мы предположили, что a, b и c — некоторые числа, причем $a \neq 0$. Однако полученные корни останутся корнями записанного выше уравнения и если рассматривать a, b и c как параметры. В этом можно убедиться простой подстановкой. Остается невыясненным лишь смысл выражения $\sqrt{b^2 - 4ac}$. Однако, как мы увидим при изучении теории поля, и это выражение имеет смысл. Таким образом, все уравнения, с которыми мы познакомились выше, можно рассматривать как *уравнения, содержащие параметры*, то есть уравнения, в левой части которых стоит *выражение, содержащее параметры*. Приведем пример, показывающий, какое важное значение имеют такие уравнения.

Зная длину a одной из сторон треугольника и длину m опущенной на нее высоты, площадь T можно вычислить из соотношения $T = \frac{1}{2}am$. Если из трех величин, T, a и m, известны две, то с помощью указанного соотношения можно определить и третью. Если, например, требуется определить высоту, то имеем: $m = \frac{2T}{a}$. Это соотношение имеет место всегда (за исключением случая $a = 0$, который из геометрических соображений все равно не представляет интереса). В каждом конкретном случае высоту можно определить подстановкой. Вообще *«параметрическими уравнениями» имеет смысл заниматься до тех пор, пока это не вызывает затруднений.* При конкретных вычислениях это позволяет избежать накопления ошибок на каждом шагу.

Пример: Найдем объем правильной четырехугольной пирамиды, если ее высота равна 10 см, а высота боковой грани 13 см. В результате последовательных вычислений, которые мы здесь опустим, получаем, что сторона основания равна 16, 6132 см, а искомый объем — 119,5993 см^3. Обозначая высоту пирамиды через m, высоту боковой грани через b и вычисляя «в параметрическом виде», получаем, что искомый объем $V = \frac{4(b^2 - m^2)m}{3}$, откуда $V = 119{,}6$ см^3. Несовпадение результатов объясняется тем, что при определении стороны основания мы извлекали корень, но смогли получить лишь приближенное значение. В то же время при определении объема нужен только квадрат стороны основания, то есть извлечение корня оказалось лишним действием.

При решении задач, приводящихся к уравнениям, зависимость между неизвестными величинами выражается, как правило, не такими уравнениями, как те, с которыми мы познакомились. Несмотря на это, необходимость в уравнениях других типов появляется очень редко. Дело в том, что получаемое уравнение часто можно «преобразовать» к одному из известных типов.

Преобразование уравнений. Рассмотрим простейшие преобразования.

Пусть, например, дано уравнение $\frac{1}{x-2} - \frac{1}{x+2} = \frac{1}{3}$. Представим себе сначала, что x — некоторое число, и выполним преобразование, не принимая во внимание никаких ограничений. Умножая обе части этого уравнения на $3(x-2)(x+2)$, приходим к уравнению $3(x+2) - 3(x-2) = x^2-4$. Отсюда после раскрытия скобок и приведения подобных членов в левой части, получаем уравнение $12 = x^2-4$. Прибавляя к обеим частям последнего уравнения 4, имеем: $x^2 = 16$. Таким образом $x = 4$ или $x = -4$. Подстановкой можно убедиться, что оба эти числа действительно являются корнями исходного уравнения.

Если же аналогичным образом преобразовать уравнение $\frac{x^2}{x-2} = \frac{5x-6}{x-2}$, то в качестве корней мы получим числа 2 и 3. Второе число действительно является корнем, но при $x = 2$ знаменатели в обеих частях уравнения обращаются в 0, то есть 2 не является корнем.

Посмотрим, какие методы следует использовать в общем случае, и какие из них можно использовать, то есть на что нужно обращать особое внимание. Однако с самого начала приходится оговориться, что общего «рецепта», который всегда приводил бы к цели, не существует. Нередко случается, что в совсем казалось бы безнадежной ситуации выручает какой-нибудь удачный прием. Однако при известном навыке удается овладеть большинством из таких приемов.

Уравнения задаются в виде $f(x) = g(x)$ (см. гл. «Теория поля»), где $f(x)$ и $g(x)$ — некоторые «выражения«. Мы рассмотрим здесь лишь частный случай, когда $f(x)$ и $g(x)$ представляют собой выражения, полученные из выражений, алгебраических относительно x, с помощью конечного числа операций сложения, вычитания, умножения и деления. Так, например, выражение $\sqrt{y} \cdot x^2 + 2^t$ мы будем считать допустимым, а выражения $\sqrt{x}$, 5^x или $\sin x$ — нет. В этом случае решить уравнение — это значит найти такое выражение, которое при подстановке вместо x превращает уравнение в тождество. (Мы используем здесь слово «выражение», так как возможно, что уравнение содержит параметры. Если же в уравнении присутствуют только числа, то понятно, что и вместо x следует подставлять число.)

Целью решения уравнения является разыскание всех его корней, так как если некоторые из них не будут найдены, то может случиться, что как раз эти потерянные корни необходимы для решения исходной задачи.

Полезно следить и за тем, чтобы в процессе решения не приобретать новых «корней», так как проверка требует дополнительных вычислений.

Если в любой из частей уравнения произвести некоторое действие, то каждый корень исходного уравнения будет корнем и полученного уравнения.

Под действием здесь понимается *сложение, вычитание или умножение двух многочленов относительно* x, а также *сложение, вычитание или умножение двух дробных алгебраических выражений относительно* x.

В самом деле, если вместо x подставить некоторое выражение в уравнение, полученное после выполнения одного из этих действий, то мы получим тот же результат, что и при аналогичной подстановке в исходное уравнение. Этим и обеспечивается возможность выполнения таких действий.

Если ($u(x)$, $v(x)$, $p(x)$ и $q(x)$ — *многочлены относительно* x, *то вместо* $\frac{u(x)}{v(x)}:\frac{p(x)}{q(x)}$ *можно писать* $\frac{u(x)}{v(x)}\cdot\frac{q(x)}{p(x)}$. И здесь результаты подстановок в исходное и получаемое выражения совпадают.

Сокращение. Дроби можно сокращать, однако при этом могут появиться новые корни. Мы будем считать допустимым лишь следующее сокращение: пусть $f(x)$, $g(x)$ и $h(x)$ — некоторые многочлены, тогда $\frac{f(x)\cdot h(x)}{g(x)\cdot h(x)}=\frac{f(x)}{g(x)}$.

Если c — такое выражение, для которого $h(c)\neq 0$, то $\frac{f(c)\,h(c)}{g(c)\,h(c)}=\frac{f(c)}{g(c)}$ (в том смысле, что если знаменатель в одной из частей равенства обращается в 0, то при этом и знаменатель в другой части обращается в 0, то есть выражения в обеих частях равенства имеют смысл или теряют его одновременно). Если же $h(c)=0$, то c может оказаться корнем нового уравнения, хотя не может быть корнем исходного уравнения, так как при подстановке получается дробь, знаменатель которой равен 0.

Очевидно, что с помощью таких преобразований мы можем привести уравнение к виду $\frac{f(x)}{g(x)}=\frac{h(x)}{k(x)}$, где $f(x)$ $g(x)$, $h(x)$ и $k(x)$ — некоторые многочлены (приведение к такому виду происходит за конечное число шагов, так как в обеих частях исходного уравнения требовалось произвести лишь конечное число операций).

Все корни уравнения $\frac{f(x)}{g(x)}=\frac{h(x)}{k(x)}$ *являются корнями и для уравнения* $f(x)\ k(x)=g(x)\ h(x)$.

Справедливость этого утверждения очевидна, так как если при некотором значении $x=c$ значения обеих дробей совпадают, то должно выполняться равенство $f(c)\ k(c)=g(c)\ h(c)$. Обратное утверждение уже неверно, так как если $g(c)=k(c)=0$, то c оказывается корнем полученного уравнения, но при подстановке в исходное уравнение лишает его смысла. Таким образом, «*обе части уравнения можно умножить на произведение знаменателей (и даже на наименьшее общее кратное знаменателей)*».

Итак, мы привели исходное уравнение к виду $f(x)=g(x)$, где в обеих частях стоят уже многочлены. Поскольку равенство $f(c)=g(c)$ выполняется тогда и только тогда, когда $f(c)-g(c)=0$, то *корни уравнения* $f(x) = g(x)$ *совпадают с корнями уравнения* $f(x)-g(x)=0$.

Это уравнение уже такого типа, как те, с которыми мы познакомились выше. Отметим еще один полезный факт:

если $f(x)$ *и* $g(x)$ *являются многочленами, то множество корней уравнения* $f(x)\,g(x)=0$ *совпадает с совокупностью корней уравнений* $f(x)=0$ *и* $g(x)=0$.

Это очевидным образом вытекает из того, что произведение равняется нулю тогда и только тогда, когда хотя бы один из сомножителей равен нулю. Отсюда следует, что *уравнение нельзя делить на многочлен от неизвестного, так как в этом случае можно «потерять» корни.*

Если, например, мы хотим решить уравнение $2x = x$ «разделив на x», то придем к равенству $2=1$, что невозможно. В то же время уравнение можно решить: действительно, $2x-x=0$, откуда $x=0$.

Рассмотренные преобразования применимы и к уравнениям, содержащим параметры, однако при выборе параметров также необходимо учитывать сделанные выше замечания.

КОМБИНАТОРИКА, ОПРЕДЕЛИТЕЛИ

ОСНОВНЫЕ ПОНЯТИЯ КОМБИНАТОРИКИ

Комбинаторика является разделом математики, независимым от алгебры. По этой причине, несмотря на значительное развитие комбинаторики в наши дни, мы познакомимся здесь лишь с основными ее задачами.

Пусть имеется n элементов (например, чисел или букв). Нас будет интересовать, каким образом можно расположить эти элементы по k местам, следующим друг за другом.

В зависимости от способа расположения мы будем различать четыре случая.

РАЗМЕЩЕНИЯ

Если считаются различными расположения, отличающиеся порядком элементов, то мы будем говорить о размещениях.

(Таким образом, при размещениях существенно, в каком порядке расположены элементы.)

Размещения без повторений. Если каждый элемент может быть помещен только на одно место, то мы имеем дело с размещениями без повторений.

Рассмотрим k «мест», по которым нужно разместить элементы:

1-е место	2-е место	...	k-е место

На первое место можно поместить любой из n элементов, то есть для заполнения первого места имеется n возможностей. Если какой-либо из данных

n элементов уже помещен на 1-е место, то на 2-е место можно поместить один из оставшихся $n-1$ элементов. Таким образом, для заполнения 2-го места имеется $n-1$ возможностей, независимо от того, который из элементов был помещен на 1-е место. Поскольку для заполнения 1-го места имелось n возможностей, и в каждом из этих n случаев 2-е место можно было заполнить $n-1$ способами, то для заполнения первых двух мест имеется $n(n-1)$ возможностей. Какие бы два элемента ни были теперь помещены на первые два места, для заполнения 3-го места следует выбрать один из оставшихся $n-2$ элементов, что можно сделать $n-2$ способами. Аналогично предыдущему, можно показать, что для заполнения первых трех мест имеется $n(n-1)(n-2)$ возможностей. Продолжая этот процесс, можно убедиться, что для заполнения всех k мест имеется $n(n-1)(n-2)\ldots(n-k+1)$ возможностей. В рассмотренном случае мы имели дело с *размещениями без повторений из n элементов по k*. Число таких размещений дается формулой

$$\boxed{A_{n,k} = n(n-1)(n-2)\ldots(n-k+1) \quad \text{(где следует предположить, что } k \leq n)}$$

В случае $n = k$ размещение без повторений называется перестановкой из n элементов. Число перестановок из n элементов обозначают через p_n.

Таким образом, это число показывает, сколькими способами можно расположить в ряд n элементов. Очевидно, что

$$\boxed{p_n = 1 \cdot 2 \cdot \ldots \cdot (n-1) \cdot n.}$$

Произведение целых чисел от 1 до n обозначается символом $n!$ (читается: *n факториал*). При $n = 1$ произведение не имеет смысла, однако в интересах общности по определению полагают $1! = 1$ (это в точности соответствует тому, что один элемент можно «расположить в ряд» одним способом). По определению полагают также, что $0! = 1$.

При помощи факториалов полученные до сих пор формулы можно записать следующим образом:

$$\boxed{p_n = n! \qquad A_{n,k} = \frac{n!}{(n-k)!}}$$

(При $n = k$ в знаменателе числа размещений без повторений появляется $0!$ и, поскольку по определению $0! = 1$, то мы как раз получаем, что $P_n = n!$).

Размещения с повторениями. Если любой из данных элементов можно поместить на несколько мест, то говорят о размещениях с повторениями.

Заполнение мест здесь может происходить аналогично тому, как это было сделано в случае размещений без повторений. Следует лишь учесть, что при заполнении каждого места могут быть использованы и элементы, уже попавшие на предыдущие места. Ввиду этого для заполнения каждого места — независимо от элементов, стоящих на предыдущих местах, — имеется n возможностей. Значит, все k мест можно заполнить n^k способами. В этом случае мы получаем *размещения с повторениями из n элементов по k.* Их число

$$\boxed{A_{n,k}^{\text{повт}} = n^k \quad (\text{здесь может быть и } k > n).}$$

Хорошим примером на размещения с повторениями является ТОТО.* Здесь нужно расположить три элемента (1, 2 и x) по 14 местам. Естественно, что любой из элементов можно поставить на несколько мест. Очевидно, что существенно, в каком порядке записываются элементы. Число всех возможностей $P_{3,14}^{\text{повт}}\, 3^{14} = 4\,782\,969$. Такое количество билетов необходимо заполнить (разными способами), чтобы обеспечить самый крупный выигрыш (то есть угадать исходы всех 14 матчей).

СОЧЕТАНИЯ

Если читаются одинаковыми расположения, отличающиеся только порядком элементов, то будем говорить о сочетаниях. (Таким образом, при сочетаниях несущественно, в каком порядке расположены элементы.)

Сочетания без повторений. Если каждый элемент может быть помещен только на одно место, то мы имеем дело с сочетаниями без повторений. *Между сочетаниями без повторений и размещениями без повторений* существует тесная связь. Предположим, что имеется n элементов и k мест. Сочетание можно однозначно определить, указав, какие из n элементов должны попасть на k мест. Как бы мы теперь ни переставляли эти k элементов местами, всякий раз будет получаться одно и то же сочетание. С другой стороны, такими перестановками можно получить все размещения из n элементов по k, причем каждое — только один раз. Таким образом, если обозначить искомое число сочетаний через $C_{n,k}$, то $C_{n,k} \cdot P_k = A_{n,k}$. Отсюда получаем формулу для числа *сочетаний из n элементов по k*:

$$\boxed{C_{n,k} = \frac{n!}{(n-k)! \cdot k!} \qquad (\text{где } k \leq n).}$$

* *Лотерея, в которой нужно предсказать исходы 14-ти футбольных матчей.* На билете имеется 14 клеток, против каждой из которых указаны участники соответствующего матча. Если предсказывается выигрыш хозяев поля, то в соответствующую клетку записывается 1, если проигрыш — 2, если ничья — х. Для получения денежного выигрыша необходимо угадать не меньше 10 исходов. (Прим. перев.)

Числа $C_{n,k}$ $\left(\text{используются также обозначения } C_n^k \text{ и } \binom{n}{k}\right)$ называются *биномиальными коэффициентами* (причина этого будет объяснена немного позже).

Сокращая на $(n-k)!$, получаем:

$$\boxed{C_{n,k} = \frac{n(n-1)(n-2)\ldots(n-k+2)(n-k+1)}{1\cdot 2\cdot 3\cdot\ldots\cdot(k-1)k}.}$$

Слева стоит целое число, следовательно, числитель в правой части делится (нацело) на знаменатель. Другими словами, *произведение любых k последовательных натуральных чисел делится на произведение первых k натуральных чисел.*

Хорошо известным примером на сочетания без повторений является ЛОТО.* Здесь из 90 чисел нужно выбрать 5. Поскольку каждое число можно выбрать только один раз, и порядок выбора чисел не имеет значения, то речь идет о сочетаниях из 90 элементов по 5. Их число $C_{90,5} = \frac{90\cdot 89\cdot 88\cdot 87\cdot 86}{1\cdot 2\cdot 3\cdot 4\cdot 5} = 43\,949\,268$. Следовательно, лишь заполнив почти 44 миллиона билетов можно быть уверенным, что — какие бы 5 чисел ни оказались выигрышными — среди заполненных билетов найдется такой, на котором зачеркнуты как раз эти 5 чисел.

Сочетания с повторениями. Если любой из элементов можно поместить на несколько мест, то говорят о сочетаниях с повторениями.

Сочетания с повторениями можно привести к сочетаниям без повторений. В случае n элементов и k мест эти сочетания называют *сочетаниями с повторениями из n элементов по k* и обозначают через $C_{n,k}^{\text{повт}}$. Их число

$$\boxed{C_{n,k}^{\text{повт}} = C_{n+k-1,k} = \frac{n(n+1)(n+2)\ldots(n+k-1)}{1\cdot 2\cdot 3\cdot\ldots\cdot k} \quad \text{(здесь также может быть } k > n)}$$

Заметим, что в формуле для числа сочетаний без повторений в числителе стоит произведение k натуральных чисел, начиная с n «вниз», в то время как для сочетаний с повторениями — начиная с n «вверх».

Некоторые свойства биномиальных коэффициентов. Учитывая, что по определению $0! = 1$, получаем: $\binom{n}{0} = 1$, перечислим теперь некоторые свойства биномиальных коэффициентов.

* Лотерея, аналогичная спортлото. Из 90 чисел нужно угадать 5. (Прим. перев.)

1. $\binom{n}{i} = \binom{n}{n-i}$ где $0 \leq i \leq n$.

2. $\binom{n+1}{k+1} = \binom{n}{k} + \binom{n}{k+1}$ где $0 \leq k \leq n$.

3. $\binom{n+1}{k+1} = \binom{n}{k} + \binom{n-1}{k} + \ldots + \binom{k+1}{k} + \binom{k}{k}$ где $0 \leq k \leq n$.

4. $\binom{n}{0} + \binom{n}{1} + \ldots + \binom{n}{n-1} + \binom{n}{n} = 2^n$.

Эти свойства можно доказать методами комбинаторики. Справедливость первых двух свойств может быть легко проверена. Выполнение третьего свойства можно доказать методом полной индукции с использованием второго свойства. О доказательстве четвертого свойства речь пойдет ниже.

Рассмотрим теперь пример, иллюстрирующий одновременно все четыре рассмотренные выше комбинаторные задачи. Сколькими способами можно раздать 6 плиток шоколада 14 лицам? Математически задача сформулирована некорректно. Следует добавить, что имеется 14 разных лиц, так как случаи, когда один и тот же шоколад получают разные лица, должны считаться различными (в действительности это совсем не одно и то же). В первый момент кажется целесообразным «рассадить» людей и раздавать шоколад, то есть считать людей «местами», а плитки шоколада «элементами». Однако в этом случае число элементов оказалось бы меньше количества мест, что привело бы к затруднениям. Ввиду этого целесообразно рассмотреть задачу наоборот: разложим плитки шоколада и будем посылать людей за ними. Итак, имеется 6 мест, по которым нужно разместить 14 элементов. Теперь, однако, выясняется, что не только математическая формулировка задачи была неточной, но и на практике такая задача неоднозначна. Действительно, может представиться несколько возможностей. Во-первых, существенно, может ли одно лицо получить несколько плиток шоколада. Во-вторых, число возможностей зависит и от того, являются ли плитки одинаковыми или различными. Мы рассмотрим здесь лишь два крайних случая, когда все плитки одинаковы или когда все они различны. Проанализируем отдельно каждый из случаев.

1. Если каждое лицо может получить только одну плитку, то это значит, что каждый человек может «попасть» только на одно место, то есть повторение элементов не допускается. Если же каждое лицо может получить несколько плиток, то каждого из них можно «поместить» на несколько мест, то есть допускается повторение элементов.

2. Если все плитки различны, то существенно, какую плитку кто получает, то есть имеет значение, какой элемент попадает на какое место. Сле-

довательно, в этом случае мы имеем дело с размещениями. Если же все плитки одинаковы, то все равно, которую из плиток получает данное лицо (важно лишь, сколько плиток оно получает). Таким образом, в этом случае не играет роли порядок элементов, то есть речь идет о сочетаниях. Итак, имеются следующие возможности:

Сколько может получить каждое лицо:		одну плитку — без повт.	несколько — с повт.
Плитки	одинаковые — сочетания	$C_{14,6}$	$C_{14,6}^{\text{повт}}$
	различные — размещения	$A_{14,6}$	$A_{14,6}^{\text{повт}}$

Имеем: $C_{14,6} = 3003$, $C_{14,6}^{\text{повт}} = 27132$, $A_{14,6} = 2\,162\,160$

и $A_{14,6}^{\text{повт}} = 7\,529\,536$.

ФОРМУЛА НЬЮТОНА

Формула Ньютона устанавливает разложение для натуральной степени, так называемого бинома, то есть выражения, представляющего собой сумму двух слагаемых. Мы знаем, что

$$(a+b)^2 = a^2 + 2ab + b^2 \qquad (a+b)^3 = a^3 + 3a^2b + 3ab + b^3.$$

Эти равенства можно записать следующим образом:

$$(a+b)^2 = \binom{2}{0} a^2 + \binom{2}{1} ab + \binom{2}{2} b^2 \quad \text{и}$$

$$(a+b)^3 = \binom{3}{0} a^3 + \binom{3}{1} a^2b + \binom{3}{2} ab^2 + \binom{3}{3} b^3.$$

Аналогичная запись возможна и в случае, если показатель равен 1 или даже 0:

$$(a+b)^1 = \binom{1}{0} a + \binom{1}{1} b \qquad (a+b)^0 = \binom{0}{0}.$$

(Последнее следует из того, что $0! = 1$). Можно доказать, что и в общем случае

$$(a+b)^n = \binom{n}{0} a^n + \binom{n}{1} a^{n-1}b + \ldots + \binom{n}{n-1} ab^{n-1} + \binom{n}{n} b^n.$$

Это равенство называется *формулой Ньютона.*

Доказательство. Мы проведем доказательство методом полной индукции. При $n = 1$ формула справедлива (больше того, как мы видели, формула верна и при $n + 0$, $n = 2$ и $n = 3$). *Предположим,* что формула справедлива для некоторого натурального числа k, и, *основываясь на этом, докажем,* что она справедлива и для $(k+1)$. Итак, по условию

$$(a+b)^k = \binom{k}{0} a^k + \binom{k}{1} a^{k-1}b + \ldots + \binom{k}{k-1} ab^{k-1} + \binom{k}{k} b^k.$$

$(k+1)$-ю степень можно разложить следующим образом:

$$(a+b)^{k+1} = (a+b)^k(a+b) = (a+b)^k a + (a+b)^k b.$$

Разложим отдельно каждое из двух слагаемых:

$$(a+b)^k a = \binom{k}{0} a^{k+1} + \binom{k}{1} a^k b + \ldots + \binom{k}{k-1} a^2 b^{k-1} + \binom{k}{k} ab^k,$$

$$(a+b)^k b = \binom{k}{0} a^k b + \binom{k}{1} a^{k-1} b + \ldots + \binom{k}{k-1} ab^k + \binom{k}{k} b^{k+1}.$$

Складывая полученные результаты, получаем:

$$(a+b)^{k+1} = \binom{k}{0} a^{k-1} + \ldots + \left[\binom{k}{i} + \binom{k}{i-1}\right] a^{k+1-i} b^i + \ldots + \binom{k}{k} b^{k+1}.$$

Поскольку $\binom{k}{0} = \binom{k}{k} = \binom{k+1}{0} = \binom{k+1}{k+1} = 1$, и по свойству 2 биномиальных коэффициентов $\binom{k}{i} + \binom{k}{i-1} = \binom{k+1}{i}$, то получается как раз требуемый результат, то есть формула доказана для $n = k+1$. Таким образом доказаны оба шага полной индукции и, следовательно, формула справедлива при любом натуральном n.

Заметим, что при $a = b = 1$ получается 4-е свойство биномиальных коэффициентов, которое тем самым также доказано.

Пример: Формулу Ньютона можно использовать для возведения двузначных чисел в степень. Чему равно, например, $24^5 = (20+4)^5$? В разложении 5-ой степени бинома появляются следующие коэффициенты: 1, 5, 10, 10, 5, 1. Отсюда: $(20+4)^5 = 20^5 + 5 \cdot 20^4 \cdot 4 + 10 \cdot 20^3 \cdot 4^2 + 10 \cdot 20^2 \cdot 4^3 + 5 \cdot 20 \cdot 4^4 + 4^5 = 3\,200\,000 + 3\,200\,000 + 1\,280\,000 + 256\,000 + 25\,600 + 1024 = 7\,962\,624$.

ПОНЯТИЕ ИНВЕРСИИ В ПЕРЕСТАНОВКЕ

Вернемся теперь к перестановкам. Рассмотрим n элементов, расположенных в некотором порядке, который мы зафиксируем. Для простоты выберем в качестве элементов числа $1, 2, \ldots, n$, расположенные в порядке возрастания. Если рассмотреть теперь некоторую перестановку этих элементов, то найдутся числа, расположенные в обратном порядке. В таком случае говорят об *инверсии*. Точнее:

если в некоторой перестановке $a_1, \ldots, a_i, \ldots, a_j, \ldots, a_n$ чисел $1, 2, \ldots, n$ выполняется неравенство $a_i > a_j$, то говорят, что числа a_i и a_j составляют инверсию.

Под числом инверсий перестановки $a_1, a_2, \ldots, a_n$ понимается количество пар элементов, составляющих инверсию (то есть число пар a_i, a_j, для которых $i < j$, но $a_i > a_j$). Перестановка называется четной, если ее элементы составляют четное число инверсий, и нечетной — в противном случае.

Число инверсий перестановки $a_1, a_2, \ldots, a_n$ обозначается через $I(a_1, a_2, \ldots, a_n)$.

Рассмотрим, например, перестановку 6, 5, 2, 3, 1, 8, 4, 7 чисел 1, 2, 3, 4, 5, 6, 7, 8. Инверсию составляют следующие пары: (6,5), (6,2), (6,3), (6,1), (6,4); (5, 2), (5,3), (5,1), (5,4); (2,1); (3,1); (8,4), (8,7). Таким образом $I(6, 5, 2, 3, 1, 8, 4, 7) = 13$.

Если числа расположены в порядке возрастания (и только в этом случае) число инверсий равно 0. Максимальное число инверсий равно числу пар, которые можно образовать из элементов перестановки; в случае n элементов это число равно $\binom{n}{2}$. Перестановка, имеющая максимальное число инверсий, существует — это перестановка, в которой элементы расположены в порядке убывания.

Если в перестановке поменять местами два элемента (такое преобразование называется транспозицией), то число инверсий, вообще говоря, изменится (ниже мы увидим, что при этом число инверсий всегда меняется). Проще всего тот случай, когда меняются местами два соседних элемента; пусть это будут a_i и a_j. Если в исходной перестановке пара a_i, a_j составляла инверсию, то после транспозиции эта инверсия пропадет и наоборот: если до транспозиции они не составляли инверсии, то после транспозиции появится новая инверсия. Во всех остальных парах порядок элементов не изменится, поэтому при транспозиции соседних элементов число инверсий в перестановке меняется на 1 (увеличивается или уменьшается). Предположим теперь, что между элементами рассматриваемой пары расположено k элементов, то есть перестановка имеет вид

$$\ldots, a, c_1, c_2, \ldots, c_k, b, \ldots$$

Элементы a и b можно поменять местами следующим образом: сначала по очереди поменяем местами a с каждым из элементов $c_1, c_2 \ldots, c_k$:

$\ldots, c_1, c_2, \ldots, c_k, a, b, \ldots$ (произведено k транспозиций соседних элементов). Поменяем теперь местами a и b:

$\ldots, c_1, c_2, \ldots, c_k, b, a, \ldots$ (еще 1 транспозиция).

И, наконец, по очереди поменяем местами b с каждым из элементов $c_k, \ldots, c_2, c_1$:

$\ldots, b, c_1, c_2, \ldots, c_k, a, \ldots$ (еще k транспозиций).

Итак, всего потребовалось произвести $2k + 1$ транспозиций соседних элементов. Поскольку при каждой такой транспозиции число инверсий перестановки изменяется на 1, то *всякая транспозиция меняет четность перестановки.*

МАТРИЦЫ И ОПРЕДЕЛИТЕЛИ

МАТРИЦЫ

В математике очень часто встречаются числа, расположенные по определенной схеме. С такой схемой мы встречались, например, при рассмотрении метода Горнера. Другим примером могут служить системы линейных уравнений, которые можно охарактеризовать при помощи коэффициентов таким образом, что по расположению коэффициентов можно установить, к какому уравнению и к какому из неизвестных относится каждый из них.

Расположение чисел в виде прямоугольной таблицы

$$\begin{bmatrix} a_{11} & a_{12} & \ldots & a_{1k} \\ a_{21} & a_{22} & \ldots & a_{2k} \\ \vdots & & & \\ a_{n1} & a_{n2} & \ldots & a_{nk} \end{bmatrix}$$

называется матрицей. Если $n = k$, то имеем квадратную матрицу порядка n.

$n \times k$ чисел, записанных выше, называются *элементами матрицы*. Для обозначения матриц мы будем использовать круглые или квадратные скобки. Иногда элементы матрицы записывают также между двумя двойными вертикальными линиями.

Под i-ой строкой записанной выше матрицы понимается матрица $(a_{i1}, a_{i2}, \ldots, a_{ik})$, под j-ым столбцом — матрица $\begin{bmatrix} a_{1j} \\ a_{2j} \\ \vdots \\ a_{nj} \end{bmatrix}$; j-ым элементом i-ой строки (или i-ым элементом j-го столбца) матрицы является число a_{ij}. Существует и более короткий способ обозначения матриц в виде $(a_{ij})_{n,k}$, где число в скобках указывает на способ образования j-го элемента i-ой строки, а числа за скобками обозначают, соответственно, число строк и столбцов матрицы. Например, $(i + j^2)_{2,3}$ обозначает матрицу

$$\begin{pmatrix} 2 & 5 & 10 \\ 3 & 6 & 11 \end{pmatrix}$$

Точно так же используются и обозначения $[a_{ij}]_{n,k}$ и $\|a_{ij}\|_{n,k}$. Для квадратной матрицы после скобок указывается только одно число: $(a_{ij})_n$ обозначает матрицу $(a_{ij})_{n,n}$. Если число строк и столбцов матрицы заведомо известно, то их обозначать не обязательно, и поэтому используется обозначение (a_{ij}).

Для квадратной матрицы совокупность чисел $a_{11}, a_{22}, \ldots, a_{nn}$, называется ее главной диагональю.

Две матрицы считаются равными, если они имеют одинаковые размеры и их соответственные элементы равны между собой: $(a_{ij})_{n,k} = (b_{ij})_{p,q}$ тогда и только тогда, когда $n = p$, $k = q$ и $a_{ij} = b_{ij}$.

Если две матрицы имеют одинаковые размеры, то их можно сложить, складывая соответственные элементы: $(a_{ij})_{nk} + (b_{ij})_{n,k} = (a_{ij} + b_{ij})_{n,k}$. Очевидно, что такое сложение коммутативно и ассоциативно. Аналогично определяется и вычитание (разностью двух матриц является матрица, каждый элемент которой равен разности соответственных элементов исходных матриц).

Всякую матрицу можно умножить на любое число согласно следующему определению: $\lambda(a_{ij}) = (\lambda a_{ij})$.

ОПРЕДЕЛИТЕЛЬ

Пусть $(a_{ij})_n$ — некоторая квадратная матрица порядка n. Определителем этой матрицы является некоторое число, которое обозначается через $|a_{ij}|_n$ (или просто $|a_{ij}|$ и вычисляется по следующему правилу. Выберем из каждой строки матрицы по одному элементу так, чтобы эти элементы были из разных столбцов (то есть получим по одному элементу и из каждого столбца). Расположим эти элементы в порядке возрастания их первых индексов, то есть индексов строки. Рассмотрим теперь произведение этих элементов, умноженное на $(-1)^I$, где I обозначает число инверсий в перестановке индексов столбцов (как чисел $1, 2, \ldots, n$). Сумма всевозможных таких произведений и даст нам определитель матрицы. Таким образом*

$$\boxed{|a_{ij}|_n = \sum_{\text{перест}} (-1)^{I(k_1, k_2, \ldots, k_n)} a_{1k_1} a_{2k_2} \ldots a_{nk_n},}$$

где $\sum\limits_{\text{перест}}$ означает, что суммирование производится по всем перестановкам $k_1, k_2, \ldots, k_n$ чисел $1, 2, \ldots, n$.

$$\text{Пример}: \begin{vmatrix} a & b \\ c & d \end{vmatrix} = (-1)^0 ad + (-1)^1 cb = ad - bc;$$

$$\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = (-1)^0 a_{11}a_{22}a_{33} + (-1)^1 a_{11}a_{23}a_{32} + (-1)^1 a_{12}a_{21}a_{33} + \\ + (-1)^2 a_{12}a_{23}a_{31} + (-1)^2 a_{13}a_{21}a_{32} + (-1)^3 a_{13}a_{22}a_{31}.$$

Определитель $|a_{ij}|_n$ называется определителем n-го порядка.

Определители играют исключительно важную роль при решени систем линейных уравнений, а вследствие этого и во многих других разделах математики и, в первую очередь, в геометрии и в математическом анализе.

Простейшие свойства определителей. Выполнение следующих свойств определителей легко проверяется.

1. *Если все элементы главной диагонали квадратной матрицы равны 1, а все остальные элементы равны 0, то определитель матрицы равен 1.*

* Используется также обозначение $\det(A_{ij})$ (прим. ред.)

2. *Если одну из строк матрицы умножить на число λ, то определитель полученной матрицы будет равен произведению определителя исходной матрицы на λ.*

3. *Предположим, что две матрицы отличаются друг от друга только i-ыми строками. Составим новую матрицу, переписав одинаковые строки без изменения, а на место i-ой строки записав сумму i-ых строк исходных матриц. Тогда определитель новой матрицы будет равен сумме определителей исходных матриц.*

Принимая во внимание, что при любой транспозиции число инверсий перестановки изменяется на нечетное число, нетрудно убедиться и в справедливости следующего свойства:

4. *Если поменять местами две строки матрицы то определитель полученной матрицы будет отличаться от определителя исходной матрицы лишь знаком.*

Используя свойства перестановок, нетрудно доказать следующее утверждение:

5. *Если элементы матрицы зеркально отразить относительно ее главной диагонали (при этом строки станут столбцами, а столбцы — строками; такое преобразование называется транспонированием матрицы), то определитель полученной матрицы будет равен определителю исходной матрицы.* Отсюда следует, что 1-е, 2-е, 3-е и 4-е свойства определителей (а также последующие свойства) останутся справедливыми, если вместо строк говорить о столбцах. Используя перечисленные свойства, можно доказать, что

если две строки матрицы равны, то ее определитель равен 0: если к одной из строк матрицы прибавить другую строку, умноженную на некоторое число, то определитель не изменится.

Это свойство дает способ вычисления определителей. Рассмотрим, например, следующий определитель:

$$\begin{vmatrix} 1 & 0 & -1 & 3 \\ 2 & 5 & 3 & -1 \\ -1 & 2 & 2 & -2 \\ -3 & 1 & 2 & 1 \end{vmatrix} = \begin{matrix}\text{Прибавляя} \\ \text{1-й столбец} \\ \text{к 3-му}\end{matrix} = \begin{vmatrix} 1 & 0 & 0 & 3 \\ 2 & 5 & 5 & -1 \\ -1 & 2 & 1 & -2 \\ -3 & 1 & -1 & 1 \end{vmatrix} = \begin{matrix}\text{Вычитая из} \\ \text{4-го столбца} \\ \text{1-й, умножен-} \\ \text{ный на 3}\end{matrix}$$

$$= \begin{vmatrix} 1 & 0 & 0 & 0 \\ 2 & 5 & 5 & -7 \\ -1 & 2 & 1 & 1 \\ -3 & 1 & -1 & 10 \end{vmatrix} = \begin{matrix}\text{Прибавляя ко 2-й строке} \\ \text{1-ю, умноженную на } -2, \\ \text{к 3-й 1-ю, к 4-й 1-ю,} \\ \text{умноженную на 3}\end{matrix} = \begin{vmatrix} 1 & 0 & 0 & 0 \\ 0 & 5 & 5 & -7 \\ 0 & 2 & 1 & 1 \\ 0 & 1 & -1 & 10 \end{vmatrix} =$$

$$\begin{matrix}\text{Вычитая 3-й столбец} \\ \text{из 4-го, а затем — 3-й,} \\ \text{умноженный} \\ \text{на 2, из 2-го}\end{matrix} = \begin{vmatrix} 1 & 0 & 0 & 0 \\ 0 & -5 & 5 & -12 \\ 0 & 0 & 1 & 0 \\ 0 & 3 & -1 & 11 \end{vmatrix} = \begin{matrix}\text{Прибавляя 3-ю строку} \\ \text{к 4-й и вычитая 3-ю,} \\ \text{умноженную на 5,} \\ \text{из 2-й}\end{matrix}$$

$$= \begin{vmatrix} 1 & 0 & 0 & 0 \\ 0 & -5 & 0 & -12 \\ 0 & 0 & 1 & 0 \\ 0 & 3 & 0 & 11 \end{vmatrix} = \begin{matrix}\text{Прибавляя}\\ \text{4-ю строку}\\ \text{ко 2-й}\end{matrix} = \begin{vmatrix} 1 & 0 & 0 & 0 \\ 0 & -2 & 0 & -1 \\ 0 & 0 & 1 & 0 \\ 0 & 3 & 0 & 11 \end{vmatrix} = \begin{matrix}\text{Вычитая из}\\ \text{2-го столбца}\\ \text{4-й, умно-}\\ \text{женный на 2}\end{matrix}$$

$$= \begin{vmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & 1 & 0 \\ 0 & -19 & 1 & 11 \end{vmatrix} = \begin{matrix}\text{Прибавляя}\\ \text{к 4-й стро-}\\ \text{ке 2-ю, умно-}\\ \text{женную на 11}\end{matrix} = \begin{vmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & 1 & 0 \\ 0 & -19 & 0 & 0 \end{vmatrix} = \begin{matrix}\text{Меняя места-}\\ \text{ми 2-ю и 4-ю}\\ \text{строки по}\\ \text{свойству 4}\end{matrix}$$

$$= - \begin{vmatrix} 1 & 0 & 0 & 0 \\ 0 & -19 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \end{vmatrix} = \begin{matrix}\text{Разделив 2-ю}\\ \text{строку на } (-19),\\ \text{а 4-ю на } (-1) \text{ по}\\ \text{свойству 2}\end{matrix} = \begin{vmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{vmatrix} =$$

$= -19$ по свойству 1.

РАЗЛОЖЕНИЕ ОПРЕДЕЛИТЕЛЯ

Часто используется другой способ вычисления определителей, однако для этого введем одно новое понятие. Если из квадратной матрицы порядка n вычеркнуть i-ю строку и j-й столбец, то мы получим некоторую квадратную матрицу порядка $n-1$. Обозначим эту матрицу через A_{ij}. Число $(-1)^{i+j}|A_{ij}|$ называется *алгебраическим дополнением* (или *дополнительной адъюнктой)* элемента a_{ij} исходной матрицы. Имеет место следующая *теорема: определитель матрицы равен сумме произведений всех элементов произвольной ее строки на их алгебраические дополнения.*

Таким образом, обозначая определитель матрицы через D, имеем:

$$\boxed{D = (-1)^{i+1}a_{i1}|A_{i1}| + (-1)^{i+2}a_{i2}|A_{i2}| + \ldots + (-1)^{i+n}a_{in}|A_{in}|\,.}$$

Заметим, что если в эту сумму вместо элементов i-ой строки записать соответственные элементы j-ой строки матрицы, то сумма окажется равной 0, то есть

$$(-1)^{i+1}a_{i1}|A_{i1}| + (-1)^{i+2}a_{j2}|A_{i2}| + \ldots + (-1)^{i+n}a_{jn}|A_{in}| = 0, \text{ если } i \neq j.$$

Присутствующие здесь знаки легко определяются по так называемому *правилу шахматной доски*. Дело в том, что знаки зависят только от расположения соответствующего элемента. При увеличении индекса строки или столбца на единицу знак меняется, то есть для элементов одной строки или одного столбца знаки меняются поочередно, как белые и черные поля на шахматной доске. Следовательно, достаточно заметить, что первому элементу первой строки соответствует знак $+$. Таким образом элементам первой строки соответствует строка знаков $+, -, +, -, +, \ldots,$

для второй строки имеем: $-, +, -, +, \ldots$, для третьей: $+, -, \ldots$, и так далее.

Рассмотрим, например, разложение по первой строке следующего определителя:

$$\begin{vmatrix} 1 & 0 & -1 & 3 \\ 2 & 5 & 3 & -1 \\ -1 & 2 & 2 & -2 \\ -3 & 1 & 2 & 1 \end{vmatrix} = (+1)1\begin{vmatrix} 5 & 3 & -1 \\ 2 & 2 & -2 \\ 1 & 2 & 1 \end{vmatrix} + (-1)0\begin{vmatrix} 2 & 3 & -1 \\ -1 & 2 & -2 \\ -3 & 2 & 1 \end{vmatrix} +$$

$$+(+1)(-1)\begin{vmatrix} 2 & 5 & -1 \\ -1 & 2 & -2 \\ -3 & 1 & 1 \end{vmatrix} + (-1)3\begin{vmatrix} 2 & 5 & 3 \\ -1 & 2 & 2 \\ -3 & 1 & 2 \end{vmatrix}$$

Теперь нужно вычислить полученные определители третьего порядка. Второй определитель нет смысла вычислять, так как один из сомножителей во втором слагаемом равен 0. Вычислив эти определители также путем разложения, мы получим определители второго порядка, которые сразу можно вычислить по формуле $\begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc$.

$$\begin{vmatrix} 5 & 3 & -1 \\ 2 & 2 & -2 \\ 1 & 2 & 1 \end{vmatrix} = 5\begin{vmatrix} 2 & -2 \\ 2 & 1 \end{vmatrix} - 3\begin{vmatrix} 2 & -2 \\ 1 & 1 \end{vmatrix} + (-1)\begin{vmatrix} 2 & 2 \\ 1 & 2 \end{vmatrix} =$$

$$= 5(2+4) - 3(2+2) - (4-2) = 30 - 12 - 2 = 16.$$

$$\begin{vmatrix} 2 & 5 & -1 \\ -1 & 2 & -2 \\ -3 & 1 & 1 \end{vmatrix} = 2\begin{vmatrix} 2 & -2 \\ 1 & 1 \end{vmatrix} - 5\begin{vmatrix} -1 & -2 \\ -3 & 1 \end{vmatrix} + (-1)\begin{vmatrix} -1 & 2 \\ -3 & 1 \end{vmatrix} =$$

$$= 2(2+2) - 5(-1-6) - (-1+6) = 8+35-5 = 38$$

$$\begin{vmatrix} 2 & 5 & 3 \\ -1 & 2 & 2 \\ -3 & 1 & 2 \end{vmatrix} = 2\begin{vmatrix} 2 & 2 \\ 1 & 2 \end{vmatrix} - 5\begin{vmatrix} -1 & 2 \\ -3 & 2 \end{vmatrix} + 3\begin{vmatrix} -1 & 2 \\ -3 & 1 \end{vmatrix} =$$

$$= 2(4-2) - 5(-2+6) + 3(-1+6) = 4 - 20 + 15 = -1.$$

Отсюда для исходного определителя получаем:

$$1 \cdot 16 - 1 \cdot 38 - 3 \cdot (-1) = 16 - 38 + 3 = -19.$$

В этом примере мы каждый раз использовали разложение по первой строке, однако можно было разлагать определители и по другим строкам. Вообще имеет смысл разлагать определитель по такой строке, в которой много нулей, так как тогда нет необходимости вычислять многие из определителей меньшего порядка. Согласно свойству 5, определитель можно разлагать не только по строке, но и по столбцу. В каждом конкретном случае следует выбирать наиболее целесообразное разложение.

ЛИНЕЙНАЯ АЛГЕБРА

ВЕКТОРНЫЕ ПРОСТРАНСТВА

Линейная алгебра возникла, по сути говоря, в процессе изучения систем линейных уравнений. Найдем, например, те числа x и y, для которых выполняются равенства

$$3x + 5y = 1 \quad \text{и} \quad 5x + 3y = 7.$$

Иными словами, нужно найти числа x и y, которые при подстановке в выражения

$$1.\ 3x + 5y - 1 \quad \text{и} \quad 2.\ 5x + 3y - 7$$

дают 0. Пусть a и b — произвольные числа, и рассмотрим выражение

$$3.\ a(3x + 5y - 1) + b(5x + 3y - 7).$$

Если числа x и y таковы, что при их подстановке выражения 1 и 2 обращаются в 0, то мы получим 0 и при подстановке этих чисел в выражение 3, так как $a \cdot 0 + b \cdot 0 = 0$. Рассмотрим, например, следующие два случая: $a = 5$, $b = -3$ и $a = 3$, $b = -5$. Имеем:

$$4.\ 5(3x + 5y - 1) - 3(5x + 3y - 7) = 16y + 16;$$

$$5.\ 3(3x + 5y - 1) - 5(5x + 3y - 7) = -16x + 32.$$

Итак, искомые числа x и y должны при подстановке в выражения $16(y + 1)$ и $-16(x-2)$ давать 0. Это условие выполняется только при $x = 2$ и $y = -1$. Действительно:

$$3 \cdot 2 + 5(-1) = 1; \quad 5 \cdot 2 + 3(-1) = 7.$$

По существу мы воспользовались тем фактом, что если выражения f_1, $f_2, \ldots, f_n$ при некоторой подстановке обращаются в 0, то при той же подстановке обращается в 0 и выражение $c_1 \cdot f_1 + c_2 \cdot f_2 + \ldots + c_n \cdot f_n$, где $c_1, c_2 \ldots, c_n$ — произвольные числа. Другими словами, рассмотрим все те выражения, которые при некоторой подстановке обращаются в 0. *Умножая такие выражения на произвольные числа и складывая их, мы всякий раз будем получать выражение такого же типа.* Суть в том, что операции сложения и умножения на число не выводят нас за пределы исходного множества математических объектов.

Этим свойством обладают очень многие типы математических объектов. Рассмотрим несколько примеров таких множеств.

1. Множество прямоугольных матриц, состоящих из n строк и k столбцов, где n и k — фиксированные натуральные числа. (Определение сложения и умножения на число см. выше).

2. Множество векторов на плоскости. (Определение сложения и умножения на число можно найти в разделе «Геометрия», стр. 288).

3. Множество векторов в пространстве.

4. Множество многочленов с вещественными коэффициентами (соответствующие определения можно найти в главе «Многочлены»).

5. Множество функций, непрерывных (или дифференцируемых, или интегрируемых и так далее) в открытом интервале (0, 1). (Под произведением функции $f(x)$ на число c понимается функция, которая в точке α принимает значение $c \cdot f(\alpha)$; под суммой двух функций $f(x)$ и $g(x)$ понимается функция, принимающая в точке α значение $f(\alpha) + g(\alpha)$).

6. Множество из предыдущего примера в случае любого другого открытого или замкнутого интервала.

Эти примеры показывают, насколько широко могут быть использованы математические системы указанного типа. Такие системы, по аналогии с множествами из 2-го и 3-го примеров, называются *векторными пространствами*. Изучая векторые пространства, можно получить результаты, справедливые в любом из перечисленных (и во многих других) конкретных случаях. В то же время, если бы мы занялись исследованием одного из приведенных конкретных примеров, то невольно воспользовались бы и такими свойствами, которые в других случаях (или в некоторых из них) не имеют места. Именно поэтому необходимо провести исследование *в общем случае*, где не используются отдельные свойства конкретных векторных пространств.

АБСТРАКТНЫЕ ВЕКТОРНЫЕ ПРОСТРАНСТВА

С этой целью вводится понятие *абстрактного векторного пространства*. Грубо говоря, абстрактным векторным пространством называется такое векторное пространство, об элементах которого — так называемых *векторах* — предполагается только, что их сумма и произведение на число также являются векторами. Несколько точнее абстрактное векторное пространство можно определить следующим образом: имеется некоторое множество элементов, природа которых нас не интересует. Складывая эти элементы, мы всякий раз получаем элемент векторного пространства. Для нас несущественно также, что означает это сложение, ведь его определение зависит от конкретного случая, наконец, умножая векторы на числа, мы опять получаем векторы; и здесь следует отвлечься от конкретного понимания умножения на число.

Однако, основываясь на сказанном до сих пор, в абстрактном векторном пространстве еще нельзя производить вычислений, так как пока неизвестно, *каким образом* (то есть с учетом каких свойств) должны производиться вычисления. Именно поэтому основные свойства операций, то есть те свойства, которые мы хотим использовать при вычислениях, нужно заранее обусловить. Иными словами, эти свойства нужно высказать как *аксиомы*. Под аксиомой обычно понимается истина, которая принимается без доказательства. В нашем случае речь идет как раз о таких истинах, так как эти свойства и не могут быть доказаны, поскольку *мы сами обусловим их выполнение*.

Прежде чем перейти к строгому определению абстрактного векторного пространства, необходимо выяснить еще один вопрос. Мы предполагаем, что векторы можно умножать на числа. Однако числа — это конкретные объекты. Для умножения на числа нам потребуется, чтобы в области чисел, которые используются, были выполнимы четыре арифметические действия, таким образом нам потребуется числовое поле.

Аксиомы абстрактного векторного пространства. Пусть Γ — некоторое числовое поле, а $\mathfrak{M}$ — некоторое множество. Назовем $\mathfrak{M}$ векторным пространством над полем Γ, если выполняются следующие аксиомы:

1. Во множестве $\mathfrak{M}$ определена операция, которую мы назовем сложением, то есть любым двум элементам $\mathbf{u}$ и $\mathbf{v}$ множества $\mathfrak{M}$ соответствует некоторый третий элемент из $\mathfrak{M}$, который обозначается через $\mathbf{u}+\mathbf{v}$.
2. Сложение обладает свойством ассоциативности, т. е. для любых элементов $\mathbf{u}$, $\mathbf{v}$ и $\mathbf{w}$ множества $\mathfrak{M}$ $(\mathbf{u}+\mathbf{v})+\mathbf{w}=\mathbf{u}+(\mathbf{v}+\mathbf{w})$.
3. Сложение обладает свойством коммутативности, т. е. для любых двух элементов $\mathbf{u}$ и $\mathbf{v}$ множества $\mathfrak{M}$ $\mathbf{u}+\mathbf{v}=\mathbf{v}+\mathbf{u}$.
4. Выполнимо вычитание, то есть для любых двух элементов $\mathbf{u}$ и $\mathbf{v}$ множества $\mathfrak{M}$ найдется такой элемент $\mathbf{x}$ из $\mathfrak{M}$, что $\mathbf{u}+\mathbf{x}=\mathbf{v}$.
5. Элементы множества $\mathfrak{M}$ можно умножать на числа из Γ, то есть любому числу α из Γ и любому элементу $\mathbf{u}$ множества $\mathfrak{M}$ соответствует некоторый элемент множества $\mathfrak{M}$, который обозначается через $\alpha\mathbf{u}$.
6. Если $\alpha\in\Gamma$ и $\mathbf{u}$, $\mathbf{v}\in\mathfrak{M}$, то $\alpha(\mathbf{u}+\mathbf{v})=\alpha\mathbf{u}+\alpha\mathbf{v}$.
7. Если α, $\beta\in\Gamma$ и $\mathbf{u}\in\mathfrak{M}$, то $(\alpha+\beta)\,\mathbf{u}=\alpha\mathbf{u}+\beta\mathbf{u}$.
8. Если α, $\beta\in\Gamma$ и $\mathbf{u}\in\mathfrak{M}$, то $(\alpha\beta)\mathbf{u}=\alpha(\beta\mathbf{u})$.
9. $1\cdot\mathbf{u}=\mathbf{u}$.

Элементы из $\mathfrak{M}$ называются векторами, а элементы поля Γ — скалярными величинами или просто скалярами. Нетрудно проверить, что в каждом из приведенных выше примеров относительно введенных операций выполняются все перечисленные аксиомы.

В первый момент кажется, что этих аксиом очень много. Однако не следует забывать, что чем меньше свойств выполняется по предположению, тем меньше можно доказать, основываясь на них. Таким образом доказать достаточно глубокие результаты можно лишь в том случае, если мы потребуем выполнения достаточно большого числа свойств. С другой стороны, большое количество аксиом трудно запомнить. Ввиду этого рассмотрим перечисленные аксиомы более подробно. Первые четыре аксиомы относятся к сложению векторов, а последующие пять — к умножению векторов на скаляры. 1-я и 5-я аксиомы обеспечивают выполнимость этих операций. 2-я, 3-я и 4-я аксиомы требуют выполнения известных свойств сложения. Собственно говоря, нужно было бы дополнительно потребовать, чтобы $\mathfrak{M}$ *было не пусто*, то есть что векторы существуют. Логически аксиомы выполнялись бы и в противном случае, но очевидно, что этот случай не представляет интереса. 6-ой, 7-ой и 8-ой аксиомами мы хотим обеспечить выполнение дистрибутивности, а также ассоциативности умножения на скаляры. Последняя аксиома как бы закрепляет то очевидное требование, что произведение любого вектора на 1 должно равняться самому вектору. Если эту аксиому отбросить, то все остальные аксиомы выполнялись бы и в таких «векторных пространствах», в которых произведение любого вектора на произвольный скаляр равнялось бы *нулевому* вектору. 9-я аксиома позволяет исключить этот случай из рассмотрения.

Основываясь на этих аксиомах, можно показать, что в любом векторном пространстве существует нулевой элемент $\mathbf{0}$, то есть такой вектор, что если $\mathbf{u}$ — произвольный вектор, то $\mathbf{0}+\mathbf{u}=\mathbf{u}$. При этом вектор $\mathbf{0}$ определен однозначно. Кроме того, можно доказать однозначность вычитания, а также то, что вектор $\alpha\mathbf{u}$ совпадает с вектором $\mathbf{0}$ тогда и только тогда, когда $\alpha=0$ или $\mathbf{u}$ совпадает с вектором $\mathbf{0}$. Можно показать также справедливость всех соотношений, которые будут использованы в дальнейшем. При изучении линейной алгебры мы будем обозначать скаляры греческими, а векторы — латинскими буквами, жирным шрифтом.

ЛИНЕЙНЫЕ КОМБИНАЦИИ ВЕКТОРОВ

При введении векторного пространства мы начали с одного из основных свойств «векторов», образовывая их «суммы с числовыми коэффициентами». Такие суммы называются линейными комбинациями векторов. Точнее: под линейными комбинациями векторов $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n$ понимаются векторы вида $\alpha_1\mathbf{u}_1+\alpha_2\mathbf{u}_2+\ldots+\alpha_n\mathbf{u}_n$, где $\alpha_1, \alpha_2, \ldots, \alpha_n$ — произвольные числа из Γ.

Если все числа α_i равны 0, то имеем *тривиальную линейную комбинацию*. Тривиальная линейная комбинация векторов всегда дает вектор $\mathbf{0}$.

Подпространство. Понятие линейной комбинации векторов тесно связано с понятием подпространства векторного пространства.

Некоторое (непустое) подмножество векторного пространства $\mathfrak{M}$ называется подпространством, если оно само является векторным пространством относительно операций, определенных в $\mathfrak{M}$.

Докажем, что

некоторое (непустое) подмножество $\mathfrak{M}_1$ векторного пространства $\mathfrak{M}$ тогда и только тогда является подпространством, когда всякая линейная комбинация любых двух векторов из $\mathfrak{M}_1$ также является вектором из $\mathfrak{M}_1$.

Для доказательства предположим сначала, что $\mathfrak{M}_1$ — подпространство пространства $\mathfrak{M}$. Пусть $\mathbf{u}, \mathbf{v} \in \mathfrak{M}_1$. Если теперь $\alpha, \beta \in \Gamma$, то по 5-ой аксиоме $\alpha\mathbf{u}, \beta\mathbf{v} \in \mathfrak{M}_1$, откуда по 1-ой аксиоме $\alpha\mathbf{u}+\beta\mathbf{v} \in \mathfrak{M}_1$.

Обратно, пусть $\mathfrak{M}_1$ — такое (непустое) подмножество пространства $\mathfrak{M}$, что если $\mathbf{u}, \mathbf{v} \in \mathfrak{M}_1$, то и $\alpha\mathbf{u}+\beta\mathbf{v} \in \mathfrak{M}_1$. Тогда в $\mathfrak{M}_1$ выполнена 1-я аксиома. Действительно, пусть $\mathbf{u}, \mathbf{v} \in \mathfrak{M}_1$. Выбирая $\alpha=\beta=1$, получаем, что $\mathbf{u}+$

$+\mathbf{v} \in \mathfrak{M}_1$. Выполнение аксиом 2 и 3 можно не проверять, так как сложение ассоциативно и коммутативно во всем векторном пространстве, а значит, и в любом его подмножестве. Для проверки 4-ой аксиомы — при данных $\mathbf{u}$, $\mathbf{v} \in \mathfrak{M}_1$ — положим $\alpha = -1$ и $\beta = 1$. Тогда — $\mathbf{u} + \mathbf{v} \in \mathfrak{M}_1$, причем $\mathbf{u} + (-\mathbf{u} + \mathbf{v}) = \mathbf{u} - \mathbf{u} + \mathbf{v} = \mathbf{0} + \mathbf{v} = \mathbf{v}$. Проверим выполнение аксиомы 6. Пусть $\mathbf{u} \in \mathfrak{M}_1$ и $\alpha \in \Gamma$. Если выбрать $\mathbf{u} = \mathbf{v}$ и $\beta = 0$, то, поскольку $\mathbf{v} \in \mathfrak{M}_1$, получаем, что $\alpha\mathbf{u} \in \mathfrak{M}_1$. Проверка остальных аксиом была бы излишней, так как они также выполняются во всем векторном пространстве. Тем самым наше утверждение доказано.

Отсюда следует, что если в векторном пространстве выбрать некоторое множество векторов (это множество может быть и бесконечным), то всевозможные линейные комбинации этих векторов образуют некоторое подпространство векторного пространства. (Рассматриваются, естественно, только линейные комбинации *конечного числа* векторов из выбранного множества; однако всякий раз берется линейная комбинация, отличная от предыдущих). В самом деле, рассмотрим две произвольные линейные комбинации выбранных векторов. Поскольку в каждой линейной комбинации присутствует лишь конечное число векторов, то в двух линейных комбинациях появится не более конечного числа векторов. Пусть это будут векторы $\mathbf{u}_1$, $\mathbf{u}_2 \ldots$, $\mathbf{u}_n$, а линейные комбинации пусть имеют вид $\alpha_1\mathbf{u}_1 + \alpha_2\mathbf{u}_2 + \ldots + \alpha_k\mathbf{u}_k$ и $\beta_1\mathbf{u}_1 + \beta_2\mathbf{u}_2 + \ldots + \beta_k\mathbf{u}_k$ (некоторые из коэффициентов могут, естественно, равняться 0). Тогда для произвольных γ, $\delta \in \Gamma$ сумма первой линейной комбинации, умноженной на γ, и второй, умноженной на δ, имеет вид: $(\gamma\alpha_1 + \delta\beta_1)\mathbf{u}_1 + (\gamma\alpha_2 + \delta\beta_2)\mathbf{u}_2 + \ldots + (\gamma\alpha_k + \delta\beta_k)\mathbf{u}_k$, то есть также является линейной комбинацией некоторых из выбранных векторов. Таким образом указанные линейные комбинации действительно образуют подпространство. Поскольку эти линейные комбинации должны принадлежать ко всякому подпространству, содержащему данные векторы, то полученное нами подпространство является наименьшим из подпространств, содержащих эти векторы.

Подпространство, образованное всевозможными линейными комбинациями некоторого множества векторов, называется подпространством, порожденным этим множеством векторов.

Если подпространство, порожденное некоторым множеством элементов векторного пространства, совпадает с самим пространством, то такое множество векторов называется системой, порождающей пространство.

Следовательно, всякая система, порождающая векторное пространство, состоит из таких векторов, что любой вектор пространства может быть представлен в виде их линейной комбинации.

Напомним, что тривиальная линейная комбинация векторов всегда дает вектор $\mathbf{0}$.

Векторы, из всех линейных комбинаций которых только тривиальная линейная комбинация равна вектору $\mathbf{0}$, называются линейно независимыми.

БАЗИС ВЕКТОРНОГО ПРОСТРАНСТВА

Базисом векторного пространства называется всякая система линейно независимых векторов, порождающая данное пространство.

Можно показать, что

некоторое множество векторов тогда и только тогда является базисом векторного пространства, когда всякий элемент этого пространства единственным образом представим в виде линейной комбинации этих векторов. Если векторное пространство имеет базис из n векторов, то оно называется n-мерным векторным пространством.

Это название вполне согласуется с геометрическими представлениями о векторном пространстве. Если, например, на плоскости выбрать два перпендикулярных единичных вектора, то всякий вектор на плоскости единственным образом представим в виде линейной комбинации этих векторов с вещественными коэффициентами. Следовательно, множество векторов на плоскости образует двухмерное векторное пространство над полем вещественных чисел. Выбирая в трехмерном пространстве три перпендикулярных друг к другу единичных вектора, аналогичным образом убеждаемся, что векторы в пространстве образуют трехмерное векторное пространство над полем вещественных чисел.

В начале этой главы мы рассмотрели множество многочленов первой степени от двух неизвестных, которые при $x = 2$ $y = -1$ обращаются в 0. Эти многочлены также образуют векторное пространство (например, над полем рациональных чисел, если рассматривать только многочлены с рациональными коэффициентами). Это пространство является двухмерным. Один из его базисов образуют, например, многочлены $x - 2$ и $y + 1$. Другим базисом являются многочлены, рассмотренные в самом начале: $3x + 5y - 1$ и $5x + 3y - 7$.

В этом примере мы указали два базиса, каждый из которых состоит из двух векторов. С аналогичным положением мы встречаемся и при рассмотрении векторов на плоскости или в трехмерном пространстве. Здесь базисом пространства является любая система из двух или, соответственно, трех векторов, перпендикулярных друг к другу. Возникает, однако, вопрос, нельзя ли, например, в трехмерном пространстве найти базис из меньшего или большего числа векторов. Оказывается, что нет — для любого векторного пространства справедливо следующее утверждение:

если векторное пространство имеет базис из n элементов, то и любой базис этого пространства состоит из n элементов.

Доказательство этого утверждения довольно длинно, поэтому мы его не приводим. Само же утверждение очень существенно, так как именно оно делает однозначным определение размерности пространства.

Итак, под базисом векторного пространства понимается любая система линейно независимых векторов, порождающая данное пространство. Таким образом, размерность пространства в некоторой степени ограничивает число элементов систем, порождающих пространство, а также число линейно независимых векторов:

в n-мерном векторном пространстве число линейно независимых векторов не может превысить n; с другой стороны, всякая система векторов, порождающая данное пространство, содержит не меньше n векторов.

Ниже мы будем заниматься только конечномерными векторными пространствами, то есть такими пространствами, в которых существует базис, состоящий из конечного числа векторов.

В качестве примера рассмотрим многочлены не выше третьей степени с рациональными коэффициентами. Они образуют векторное пространство над полем рациональных чисел, если сложение определить как сложение многочленов, а умножение — как умножение многочленов на число. Подпространствами этого векторного пространства являются, например, множества многочленов не выше второй или не выше первой степени, а также множество многочленов нулевой степени. Действительно, при сложении или умножении на число степень многочленов не может увеличиться. Однако многочлены только третьей степени не образуют подпространства, так как сумма многочленов третьей степени не обязательно имеет степень 3: $(x^3 + 4x - 1) + (-x^3 + 2x^2 + 3) = 2x^2 + 4x + 2$.

Посмотрим, из каких многочленов состоит подпространство, порожденное многочленами $3x + 5$ и $2x - 1$. В этом подпространстве содержится многочлен $(3x + 5) + 5\cdot(2x - 1) = 13x$, а следовательно и многочлен $\frac{1}{13}\cdot 13x = x$. К нему должен принадлежать и многочлен $2\cdot x + (-1)(2x - 1) = 1$. Следовательно, к этому подпространству принадлежат все многочлены вида $a\cdot x + b$, то есть все многочлены, степень которых не выше единицы. Поскольку с помощью операций сложения и умножения на число нельзя получить многочлен более высокой степени, то рассматриваемое подпространство состоит из всех многочленов не выше первой степени.

Очевидно, что многочлены $1, x, x^2, x^3, \ldots$ образуют систему, порождающую все пространство, так как любой многочлен представим в виде их линейной комбинации. Больше того, эти многочлены образуют базис пространства, так как линейно независимы. Действительно, любая их линейная комбинация вида $ax^3 + bx^2 + cx + d$ может быть равна 0 только если $a = b = c = d = 0$ (речь идет о равенстве многочленов, а не об уравнениях!). Точно так же линейно независимы и указанные выше многочлены $3x + 5$ и $2x - 1$. В то же время многочлены $x^2 + 2x - 3$, $x^2 - 5x + 4$ и $3x^2 + 5x - 8$ уже не являются линейно независимыми, так как

$$\begin{aligned}&20(x^2 + 2x - 3) + 1(x^2 - 5x + 4) + (-7)(3x^2 + 5x - 8) =\\&= (20 + 1 - 7\cdot 3)x^2 + (20\cdot 2 - 5 - 7\cdot 5)x + (20\cdot(-3) + 4 +\\&+ (-7)(-8)) = 0,\end{aligned}$$

и это не тривиальная линейная комбинация. Заметим, что это векторное пространство (многочлены не выше третьей степени с рациональными коэффициентами) является четырехмерным, так как имеет базис, состоящий из четырех элементов.

ЛИНЕЙНЫЕ ОТОБРАЖЕНИЯ

В геометрии исключительно важную роль играют так называемые аффинные преобразования. На плоскости такими преобразованиями являются, например, сдвиг, поворот, растяжение и другие. Аналогичные преобразования возможны и в трехмерном пространстве. Аффинным преобразованием является и проектирование пространства на некоторую плоскость. Все эти преобразования обладают тем свойством, что сумму векторов они переводят в сумму их образов, а произведение вектора на число α — в произведение образа вектора на α. В пространстве многочленов или дифференцируемых функций аналогичным свойством обладает дифференцирование. Поскольку производная дифференцируемой функции сама не обязана быть дифференцируемой, то в последнем примере получается, что образы принадлежат к некоторому другому пространству. Обычно такие отображения называют *линейными отображениями* (или, точнее, *однородными линейными отображениями*). Дадим теперь строгое определение.

Пусть $\mathfrak{M}_1$ и $\mathfrak{M}_2$ — векторные пространства над числовым полем Γ. Поставим в соответствие каждому вектору $\mathbf{u}$ пространства $\mathfrak{M}_1$ вектор из пространства $\mathfrak{M}_2$, который обозначим через $A\mathbf{u}$. Если для любых $\alpha, \beta \in \Gamma$ $A(\alpha\mathbf{u} + \beta\mathbf{v}) = \alpha A\mathbf{u} + \beta A\mathbf{v}$, то говорят, что A есть однородное линейное отображение векторного пространства $\mathfrak{M}_1$ в векторное пространство $\mathfrak{M}_2$ и обозначают: $A\colon\ \mathfrak{M}_1 \to \mathfrak{M}_2$.

Линейные отображения $A\colon\ \mathfrak{M}_1 \to \mathfrak{M}_2$ и $B\colon\ \mathfrak{M}_1 \to \mathfrak{M}_2$ мы назовем равными, если для любого вектора $\mathbf{u}$ из $\mathfrak{M}_1$ $A\mathbf{u} = B\mathbf{u}$.

Если рассматривать линейные отображения с точки зрения теории функций, то их следует считать такими функциями, для которых одно из пространств является областью определения, а другое — областью значений. Естественно, что это функции специального типа. Теория функций занимается в первую очередь такими функциями, для которых и областью определения, и областью значений является множество вещественных чисел. Посмотрим, какие функции можно назвать в этом случае линейными отображениями. Множество вещественных чисел можно рассматривать как векторное пространство над полем вещественных чисел, базисом которого является число 1 (то есть одномерное векторное пространство). Действительно, всякое вещественное число α может быть единственным образом записано в виде $\alpha = \alpha \cdot 1$. Предположим, что линейное отображение A переводит число 1 в число $\alpha (= \alpha \cdot 1)$. Тогда образ числа ξ при отображении A можно получить так: $A\xi = A(\xi \cdot 1) = \xi \cdot A(1) = \xi \cdot \alpha$. Таким образом в этом случае мы получаем функции, которые вещественному числу ξ ставят в соответствие вещественное число $\alpha \cdot \xi$, где α — некоторое произвольное, но фиксированное число. Геометрически такая функция изображается в виде прямой, то есть это действительно «линейная» функция или линейное отображение. Нетрудно установить также, что изображением такой функции является не произвольная прямая, а прямая, проходящая через начало координат. На это обстоятельство и указывает слово «однородный» в названии линейных отображений. Очевидно, что всякое вещественное число может быть записано в виде $\alpha \cdot \xi$, если только α не равно 0. Другими словами, образами вещественных чисел при рассматриваемом отображении оказываются или все вещественные числа, или же только число 0.

То же самое верно и в случае произвольного линейного отображения $A:\ \mathfrak{M}_1 \to \mathfrak{M}_2$. Точнее:

при данном линейном отображении $A:\ \mathfrak{M}_1 \to \mathfrak{M}_2$ *элементы вида* $A\mathbf{u}$ *пространства* $\mathfrak{M}_2$ *образуют некоторое подпространство этого пространства, называемое пространством образов при отображении* A *и обозначаемое символом* $\mathrm{Im}(A)$ (Im — сокращение латинского слова imago, что означает «образ»).

Нужно лишь доказать, что элементы указанного вида действительно образуют подпространство. Однако это очевидно, так как если $\alpha, \beta \in \Gamma$, то вектор $\alpha A\mathbf{u} + \beta A\mathbf{v} = A(\alpha\mathbf{u} + \beta\mathbf{v})$ принадлежит $\mathrm{Im}(A)$ по определению.

Размерность пространства образов при отображении A называют рангом A и обозначают через $r(A)$.

ОПЕРАЦИИ НАД ЛИНЕЙНЫМИ ОТОБРАЖЕНИЯМИ

Над линейными отображениями можно производить некоторые операции: их можно умножать на число и иногда можно складывать (одни можно, другие нет). Эти операции производятся точно так же, как операции над функциями. В некоторых случаях линейные отображения можно перемножать; однако это соответствует не умножению функций, а замене переменной. Сформулируем определения этих операций.

Если $A:\ \mathfrak{M}_1 \to \mathfrak{M}_2$ *и* $\lambda \in \Gamma$, *то* λA *обозначает линейное отображение, при котором для любого* $\mathbf{u} \in \mathfrak{M}_1$ $(\lambda A)\mathbf{u} = \lambda(A\mathbf{u})$.

Нетрудно показать, что λA действительно является линейным отображением и $\lambda A:\ \mathfrak{M}_1 \to \mathfrak{M}_2$.

Если $A:\ \mathfrak{M}_1 \to \mathfrak{M}_2$ *и* $B:\ \mathfrak{M}_1 \to \mathfrak{M}_2$, *то* $A + B$ *обозначает линейное отображение, при котором для любого* $\mathbf{u} \in \mathfrak{M}_1$ $(A + B)\mathbf{u} = A\mathbf{u} + B\mathbf{u}$.

Аналогично предыдущему легко убедиться, что $A + B$ является линейным отображением и $A + B:\ \mathfrak{M}_1 \to \mathfrak{M}_2$.

Если $A:\ \mathfrak{M}_1 \to \mathfrak{M}_2$ *и* $B:\ \mathfrak{M}_2 \to \mathfrak{M}_3$, *то* BA *обозначает линейное отображение, при котором для любого* $\mathbf{u} \in \mathfrak{M}_1$ $(BA)\mathbf{u} = B(A\mathbf{u})$.

Нетрудно показать, что BA также является линейным отображением. Если $\mathbf{u} \in \mathfrak{M}_1$, то $A\mathbf{u} \in \mathfrak{M}_2$ и, поскольку $B:\ \mathfrak{M}_2 \to \mathfrak{M}_3$, то $B(A\mathbf{u}) \in \mathfrak{M}_3$. Таким образом отображение BA переводит векторы из $\mathfrak{M}_1$ в векторы из $\mathfrak{M}_3$, то есть $BA:\ \mathfrak{M}_1 \to \mathfrak{M}_3$.

С помощью несложных выкладок можно показать, что *линейные отображения векторного пространства* $\mathfrak{M}_1$ *в векторное пространство* $\mathfrak{M}_2$ *относительно определенных выше операций сложения и умножения на число сами образуют векторное пространство над полем* Γ.

(Для доказательства достаточно лишь проверить выполнение аксиом.)

Очевидно, что умножение линейных отображений не коммутативно. В то же время ассоциативность умножения выполняется (при условии, что само умножение выполнимо), то есть *если существуют произведения* CB *и* BA, *то существуют* и *произведения* $C(BA)$ и $(CB)A$, *причем* $C(BA) = (CB)A$.

Пусть $A\colon\ \mathfrak{M}_1 \to \mathfrak{M}_2$. Поскольку существует произведение BA, то B отображает $\mathfrak{M}_2$ в некоторое другое пространство, скажем в $\mathfrak{M}_3$, то есть $B\colon\ \mathfrak{M}_2 \to \mathfrak{M}_3$. Из существования произведения CB следует, что C должно отображать пространство $\mathfrak{M}_3$ в некоторое пространство $\mathfrak{M}_4$, то есть $C\colon$ $\mathfrak{M}_3 \to \mathfrak{M}_4$. Отсюда следует, что $BA\colon\ \mathfrak{M}_1 \to \mathfrak{M}_3$ и $CB\colon\ \mathfrak{M}_2 \to \mathfrak{M}_4$. Но тогда существуют и произведения $C(BA)$ и $(CB)A$, каждое из которых отображает пространство $\mathfrak{M}_1$ в пространство $\mathfrak{M}_4$. Кроме того, по определению умножения для любого элемента $\mathbf{u}$ пространства $\mathfrak{M}_1$ имеем:

$$[C(BA)]\mathbf{u} = C[(BA)\mathbf{u}] = C[B(A\mathbf{u})];$$

$$[(CB)A]\mathbf{u} = (CB)[A\mathbf{u}] = C(B[A\mathbf{u}]).$$

Следовательно, и вектор $[C(BA)]\mathbf{u}$, и вектор $[(CB)A]\mathbf{u}$ получаются таким образом, что сначала производится отображение A вектора $\mathbf{u}$, затем — отображение B вектора $A\mathbf{u}$ и, наконец, отображение C вектора $B(A\mathbf{u})$. Таким образом и $C(BA)$, и $(CB)A$ отображают каждый вектор $\mathbf{u}$ пространства $\mathfrak{M}_1$ в один и тот же вектор, то есть по определению равенства двух *отображений имеем*: $C(BA) = (CB)A$.

ЛИНЕЙНЫЕ ПРЕОБРАЗОВАНИЯ

Особенно важное значение имеют *линейные отображения векторного пространства в себя*, получившие название *линейных преобразований*.

Пусть $\mathfrak{M}$ — некоторое n-мерное векторное пространство. Для линейных преобразований пространства $\mathfrak{M}$ справедливы следующие утверждения:

1. Для любых двух линейных преобразований существуют их сумма и произведение.

2. Имеет место дистрибутивность, то есть если A, B и C — три линейные преобразования, то $(A + B)C = AC + BC$ и $C(A + B) = CA + CB$ (поскольку умножение некоммутативно, то приходится говорить о двух типах дистрибутивности).

3. Обозначим через E линейное преобразование, при котором $E\mathbf{u} = \mathbf{u}$ для любого вектора $\mathbf{u}$. Тогда для любого линейного преобразования A имеет место равенство $EA = AE = A$.

Если пространством образов преобразования A является все пространство, то есть если $r(A) = n$, то преобразование A называется *невырожденным*. Если пространство образов преобразования B не есть все пространство (то есть если $r(B) < n$), то преобразование B называется *вырожденным*.

4. Если A — невырожденное линейное преобразование, то для него найдется такое линейное преобразование A^{-1}, называемое *обратным* преобразованию A, для которого $A(A^{-1}) = (A^{-1})A = E$. Если при этом C — такое преобразование, что $CA = 0$, то $C = 0$ (то есть нулевое линейное преобразование, которое всякий вектор отображает в вектор $\mathbf{0}$).

5. Если B — вырожденное линейное преобразование, то для него не существует обратного, и найдется такое преобразование D, отличное от нулевого, для которого $DA = AD = 0$.

ЗАДАНИЕ ЛИНЕЙНЫХ ОТОБРАЖЕНИЙ С ПОМОЩЬЮ МАТРИЦ

Пусть $\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n$ — некоторый базис n-мерного пространства $\mathfrak{M}_1$. Тогда любой вектор $\mathbf{u}$ пространства единственным образом представим в виде

$$\mathbf{u} = \alpha_1\mathbf{e}_1 + \alpha_2\mathbf{e}_2 + \ldots + \alpha_n\mathbf{e}_n, \quad \text{где} \quad \alpha_1, \alpha_2, \ldots, \alpha_n \in \Gamma.$$

Согласно этому, вектору $\mathbf{u}$ можно однозначно поставить в соответствие матрицу $\begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{bmatrix}$. Элементы этой матрицы называются компонентами вектора $\mathbf{u}$. Матрицу, соответствующую вектору $\mathbf{u}$, мы будем обозначать через $[\mathbf{u}]$. Аналогичным образом с помощью матриц можно задавать и линейные отображения.

Пусть, кроме пространства $\mathfrak{M}_1$, задано некоторое пространство $\mathfrak{M}_2$, базис которого обозначим через $\mathbf{f}_1, \mathbf{f}_2, \ldots, \mathbf{f}_k$. Линейное отображение $A: \mathfrak{M}_1 \to \mathfrak{M}_2$ можно однозначно определить заданием образов элементов базиса при этом отображении. Эти образы однозначно представляются в виде линейных комбинаций векторов базиса пространства $\mathfrak{M}_2$:

$$\begin{aligned} A\mathbf{e}_1 &= \alpha_{11}\mathbf{f}_1 + \alpha_{21}\mathbf{f}_2 + \ldots + \alpha_{k1}\mathbf{f}_k \\ A\mathbf{e}_2 &= \alpha_{12}\mathbf{f}_1 + \alpha_{22}\mathbf{f}_2 + \ldots + \alpha_{k2}\mathbf{f}_k \\ &\vdots \\ A\mathbf{e}_n &= \alpha_{1n}\mathbf{f}_1 + \alpha_{2n}\mathbf{f}_2 + \ldots + \alpha_{kn}\mathbf{f}_k. \end{aligned}$$

Отсюда сразу видно, что линейное отображение A однозначно определяется этими $n \cdot k$ числами (если принять во внимание и их расположение). Другими словами, линейное отображение A может быть однозначно задано следующей матрицей:

$$[A] = \begin{bmatrix} \alpha_{11} & \alpha_{12} & \ldots & \alpha_{1n} \\ \alpha_{21} & \alpha_{22} & \ldots & \alpha_{2n} \\ \vdots & & & \\ \alpha_{k1} & \alpha_{k2} & \ldots & \alpha_{kn} \end{bmatrix}.$$

Точно так же, как в случае векторов, здесь компоненты отдельных векторов попали в соответственные столбцы. Таким образом компоненты вектора $A\mathbf{e}_1$ являются элементами первого, компоненты $A\mathbf{e}_2$ — элементами второго, ..., а компоненты вектора $A\mathbf{e}_n$ — элементами n-го столбца матрицы.

Мы уже познакомились с умножением матрицы на число и со сложением матриц. Нетрудно проверить, что

1. $[\lambda A] = \lambda[A]$;
2. если существует отображение $A + B$, то $[A + B] = [A] + [B]$.

Умножение матриц. Однако линейные отображения можно и перемножать. С помощью этого можно определить и умножение матриц.

3. Если существует отображение BA, то пусть

$$[B] \cdot [A] = [BA].$$

Нетрудно показать, что для любой прямоугольной матрицы найдется линейное отображение, которое задается этой матрицей. Основываясь на этом, выясним, когда и каким образом можно перемножить две матрицы.

Две матрицы можно перемножить тогда и только тогда, когда число столбцов в первом сомножителе равно числу строк во втором сомножителе; в этом случае число строк матрицы, являющейся произведением, будет равно числу строк в первом сомножителе, а число столбцов — числу столбцов во втором сомножителе.

Покажем, как образуются элементы матрицы, получающейся в произведении. Пусть

$$\begin{bmatrix} \alpha_{11} & \alpha_{12} & \cdots & \alpha_{1n} \\ \alpha_{21} & \alpha_{22} & \cdots & \alpha_{2n} \\ \vdots & & & \\ \alpha_{k1} & \alpha_{k2} & \cdots & \alpha_{kn} \end{bmatrix} \cdot \begin{bmatrix} \beta_{11} & \beta_{12} & \cdots & \beta_{1m} \\ \beta_{21} & \beta_{22} & \cdots & \beta_{2m} \\ \vdots & & & \\ \beta_{n1} & \beta_{n2} & \cdots & \beta_{nm} \end{bmatrix} = \begin{bmatrix} \gamma_{11} & \gamma_{12} & \cdots & \gamma_{1m} \\ \gamma_{21} & \gamma_{22} & \cdots & \gamma_{2m} \\ \vdots & & & \\ \gamma_{k1} & \gamma_{k2} & \cdots & \gamma_{km} \end{bmatrix},$$

Тогда:

$$\boxed{\gamma_{ij} = \alpha_{i1} \cdot \beta_{1j} + \alpha_{i2} \cdot \beta_{2j} + \ldots + \alpha_{in} \cdot \beta_{nj} .}$$

Нетрудно проверить также, что если известны компоненты вектора и матрица отображения, то компоненты образа вектора можно получить из соотношения

$$\boxed{[A\mathbf{u}] = [A] \cdot [\mathbf{u}] .}$$

Можно показать, что все элементы главной диагонали матрицы линейного преобразования E равны 1, а все остальные ее элементы — нулю. В то же время все элементы матрицы нулевого преобразования равны нулю, точно так же, как и компоненты вектора $\mathbf{0}$.

Очевидно, что свойства операций над отображениями выполняются и для операций над матрицами. Например, умножение матриц ассоциативно, а также дистрибутивно относительно сложения.

Ранг матрицы. Ранг матрицы можно определить следующим образом:

$$r([A]) = r(A) .$$

Можно доказать, что *ранг матрицы не изменится, если к одной из ее строк (столбцов) прибавить другую строку (столбец), умноженную на некоторое число.*

Можно показать также, что *линейное преобразование является невырожденным (то есть существует обратное ему) тогда и только тогда, когда определитель его матрицы отличен от* 0. Матрицу обратного преобразования можно получить следующим образом:

$$\text{пусть} \quad [A] = (\alpha_{ij})_n; \quad [A^{-1}] = (\beta_{ij})_n;$$

$$\text{тогда} \quad \beta_{ij} = \frac{1}{D} \cdot (-1)^{i+j} A_{ji}, \qquad \text{где } (-1)^{i+j} A_{ji}$$

является алгебраическим дополнением к элементу, стоящему на i-ом месте в j-й строке исходной матрицы, а D — определитель исходной матрицы. (Важно отметить, что в i-й строке на j-ом месте обратной матрицы стоит алгебраическое дополнение не к элементу α_{ij}, а к элементу α_{ji} исходной матрицы.)

В заключение отметим, что *определитель матрицы* $[BA]$ *равен произведению определителей матриц* $[B]$ и $[A]$.

ЕВКЛИДОВО ПРОСТРАНСТВО

Скалярное произведение векторов. Пусть $\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n$ — базис некоторого векторного пространства. Рассмотрим два произвольных вектора пространства и запишем их в виде линейных комбинаций векторов базиса:

$$\mathbf{u} = \xi_1\mathbf{e}_1 + \xi_2\mathbf{e}_2 + \ldots + \xi_n\mathbf{e}_n; \qquad \mathbf{v} = \eta_1\mathbf{e}_1 + \eta_2\mathbf{e}_2 + \ldots + \eta_n\mathbf{e}_n.$$

Под скалярным произведением векторов $\mathbf{u}$ и $\mathbf{v}$ понимается число

$$(\mathbf{u}, \mathbf{v}) = \xi_1\eta_1 + \xi_2\eta_2 + \ldots + \xi_n\eta_n.$$

Как мы видели, и при умножении матриц появляется выражение такого вида: j-й элемент i-й строки произведения двух матриц может рассматриваться как скалярное произведение i-й строки первого сомножителя на j-й столбец второго.

Если в векторном пространстве ввести скалярное произведение, то мы получим евклидово пространство. (Данное выше определение совпадает с определяемым в геометрии скалярным произведением векторов на плоскости или в трехмерном пространстве.)

Пример: скалярное произведение векторов $\mathbf{u} = 3\mathbf{e}_1 - 2\mathbf{e}_2 - 5\mathbf{e}_3 + 2\mathbf{e}_4$ и $\mathbf{v} = 7\mathbf{e}_1 + 4\mathbf{e}_2 - 2\mathbf{e}_3 - 6\mathbf{e}_4$ равно $(\mathbf{u}, \mathbf{v}) = 3 \cdot 7 - 2 \cdot 4 + 5 \cdot 2 - 2 \cdot 6 = 21 - 8 + 10 - 12 = 11$; для $\mathbf{u} = 3\mathbf{e}_1 + 2\mathbf{e}_2 - 5\mathbf{e}_3 + \mathbf{e}_4$ и $\mathbf{v} = 7\mathbf{e}_1 - 8\mathbf{e}_2 + 2\mathbf{e}_3 + 5\mathbf{e}_4$ имеем: $(\mathbf{u}, \mathbf{v}) = 21 - 16 - 10 + 5 = 0$.

Если скалярное произведение двух векторов равно 0, то векторы называются **ортогональными** (перпендикулярными друг другу). (Это также совпадает с геометрическим определением.)

Нетрудно показать, что

$$\boxed{(\mathbf{u}, \mathbf{v}) = (\mathbf{v}, \mathbf{u})\,; \quad (\alpha\mathbf{u}_1 + \beta\mathbf{u}_2, \mathbf{v}) = \alpha(\mathbf{u}_1, \mathbf{v}) + \beta(\mathbf{u}_2, \mathbf{v})\,.}$$

Рассмотрим пример, показывающий, какую важность имеет понятие скалярного произведения. Функции, непрерывные в замкнутом интервале $[a;\ b]$, образуют векторное пространство над полем вещественных чисел. Поставим в соответствие любым двум элементам $f(x)$ и $g(x)$ этого пространства число, определяемое интегралом $\int\limits_a^b f(x)\,g(x)dx$. При подходящем выборе базиса этот интеграл можно рассматривать как скалярное произведение. Рассмотрим, например, уже упоминавшееся пространство многочленов не выше третьей степени, но не только с рациональными, а с любыми вещественными коэффициентами, определенных в интервале $[-1, +1]$. Можно показать, что базис

$$f_1 = \frac{1}{2}, \quad f_2 = \sqrt{\frac{3}{2}}\cdot x, \quad f_3 = \sqrt{\frac{5}{8}}\cdot(3x^2 - 1), \quad f_4 = \sqrt{\frac{7}{8}}\cdot(5x^3 - 3x)$$

обладает соответствующим свойством, то есть если два многочлена не выше третьей степени записать в виде

$$g = \alpha_1 f_1 + \alpha_2 f_2 + \alpha_3 f_3 + \alpha_4 f_4\,; \quad h = \beta_1 f_1 + \beta_2 f_2 + \beta_3 f_3 + \beta_4 f_4\,,$$

то

$$\int\limits_{-1}^{+1} g\cdot h dx = \alpha_1\beta_1 + \alpha_2\beta_2 + \alpha_3\beta_3 + \alpha_4\beta_4\,.$$

КВАДРАТИЧНЫЕ ФОРМЫ

Во многих разделах математики и, в первую очередь, в геометрии часто встречаются суммы, в которых слагаемые являются выражениями второй степени от неизвестных (или переменных).

Итак, в общем случае речь идет о выражении вида

$$\alpha_{11}\xi_1^2 + 2\alpha_{12}\xi_1\xi_2 + \alpha_{22}\xi_2^2 + \ldots + 2\alpha_{1n}\xi_1\xi_n + 2\alpha_{2n}\xi_2\xi_n + \ldots + \alpha_{nn}\xi_n^2\,,$$

которое называется квадратичной формой от неизвестных $\xi_1, \xi_2, \ldots, \xi_n$.

Изучение квадратичных форм значительно облегчается, если использовать евклидово пространство.

Пусть $\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n$ — базис некоторого евклидова пространства (с определенным выше скалярным произведением). Рассмотрим в этом пространстве некоторый вектор $\mathbf{x} = \xi_1\mathbf{e}_1 + \xi_2\mathbf{e}_2 + \ldots + \xi_n\mathbf{e}_n$ (который будем считать «переменным» вектором). Пусть $[A] = (\alpha_{ij})_n$ — матрица линейного преобразования A, где числа α_{ij} — коэффициенты записанной выше квадратичной формы. Однако в квадратичной форме присутствуют лишь коэффициенты, индексы которых удовлетворяют неравенству $i \leq j$. Чтобы это не служило ограничением, положим по определению $\alpha_{ij} = \alpha_{ji}$. По этой причине в квадратичной форме и были записаны члены вида $2\alpha_{ij}\xi_i\xi_j$. «Половина» этого числа записывается в матрице «выше» главной диагонали, а вторая «половина» — симметрично относительно главной диагонали «снизу». После этого нетрудно показать, что рассматриваемая квадратичная форма совпадает со скалярным произведением $(A\mathbf{x}, \mathbf{x})$. Рассмотрим, например, следующие три квадратичные формы:

a) $2\xi^2 + 6\xi\eta + 5\eta^2$;

b) $\xi^2 + 4\xi\eta + 4\eta^2$,

c) $\xi^2 + 4\xi\eta + \eta^2$.

Матрицы соответствующих линейных преобразований:

$$[A] = \begin{pmatrix} 2 & 3 \\ 3 & 5 \end{pmatrix}; \qquad [B] = \begin{pmatrix} 1 & 2 \\ 2 & 4 \end{pmatrix}; \qquad [C] = \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix}$$

Если теперь $\mathbf{e}, \mathbf{f}$ — некоторый базис пространства, и $\mathbf{x} = \xi\mathbf{e} + \eta\mathbf{f}$, то соответствующие квадратичные формы можно записать следующим образом:

a) $(A\mathbf{x}, \mathbf{x})$, b) $(B\mathbf{x}, \mathbf{x})$, c) $(C\mathbf{x}, \mathbf{x})$.

Приведение к сумме квадратов. Рассмотрим некоторую квадратичную форму, которую запишем в виде $(A\mathbf{x}, \mathbf{x})$, где $\mathbf{x}$ — переменный вектор, «пробегающий» по всему n-мерному евклидову пространству $\mathfrak{M}$. Можно доказать, что в пространстве $\mathfrak{M}$ найдется такой ортонормированный базис $\mathbf{f}_1, \mathbf{f}_2, \ldots, \mathbf{f}_n$ (это означает, что элементы базиса попарно ортогональны, то есть $(\mathbf{f}_i, \mathbf{f}_j) = 0$, при $i \neq j$, и «длина» каждого элемента равна 1, то есть $(\mathbf{f}_i, \mathbf{f}_i) = 1$), что для каждого $\mathbf{x} = \eta_1\mathbf{f}_1 + \eta_2\mathbf{f}_2 + \ldots + \eta_n\mathbf{f}_n$:

$$\boxed{(A\mathbf{x}, \mathbf{x}) = \lambda_1\eta_1^2 + \lambda_2\eta_2^2 + \ldots + \lambda_n\eta_n^2,}$$

где коэффициенты λ_i — некоторые числа.

Это *основная теорема о квадратичных формах*, которая утверждает, что *всякая квадратичная форма приводится к сумме квадратов.*

Было бы сложно здесь даже задать соответствующий базис, поэтому в качестве примера мы лишь укажем, какой вид примут три рассмотренные выше квадратичные формы после приведения их к сумме квадратов:

a) $(\xi + \eta)^2 + (\xi + 2\eta)^2$.

b) $(\xi + 2\eta)^2 + 0^2$,

c) $(\xi + 2\eta)^2 + (-3)\eta^2$.

Пусть Γ — поле вещественных чисел. Квадратичная форма $(A\mathbf{x}, \mathbf{x})$ (пример а) может равняться 0 только если $\xi + \eta = \xi + 2\eta = 0$, что возможно лишь при $\xi = \eta = 0$, во всех других случаях эта квадратичная форма положительна. Квадратичная форма $(B\mathbf{x}, \mathbf{x})$ (пример b) также не может быть отрицательной, однако, например, при $\xi = 2$, $\eta = -1$ обращается в 0. Наконец, квадратичная форма $(C\mathbf{x}, \mathbf{x})$ (пример с) может быть как положительной, так и отрицательной. Если, например, $\mathbf{x} = \mathbf{e}$, то $(C\mathbf{x}; \mathbf{x}) = 1$, а если $\mathbf{x} = -2\mathbf{e} + \mathbf{f}$, то $(C\mathbf{x}, \mathbf{x}) = -3$. Соответственно этим трем случаям квадратичные формы классифицируют следующим образом:

а) если квадратичная форма всегда положительна (отрицательна), за исключением того случая, когда все переменные равны 0, то она называется положительно (отрицательно) определенной;

б) если квадратичная форма принимает лишь неотрицательные (неположительные) значения, то она называется положительно (отрицательно) полуопределенной; в) если квадратичная форма может принимать как положительные, так и отрицательные значения, то она называется неопределенной.

ХАРАКТЕРИСТИЧЕСКИЙ МНОГОЧЛЕН

В самых разных областях математики при решении численных задач очень часто возникает проблема отыскания для данного линейного преобразования A такого вектора $\mathbf{x}$, для которого $A\mathbf{x} = \lambda\mathbf{x}$, где λ — некоторое число. Такой вектор $\mathbf{x}$ можно найти не для всех чисел λ, поэтому для решения задачи необходимо сначала установить, для каких чисел λ это возможно.

Если для некоторого числа λ и вектора $\mathbf{x}$ выполняется равенство $A\mathbf{x} = \lambda\mathbf{x}$, то λ называется собственным значением преобразования A, а вектор $\mathbf{x}$ — собственным вектором, относящимся к собственному значению λ.

Таким образом задача состоит в отыскании собственных значений и собственных векторов данного преобразования A. Очень часто достаточно бывает определить только собственные значения. Этой задачей мы и займемся.

Определение собственных значений. Если $A\mathbf{x} = \lambda\mathbf{x}$, то, поскольку $E\mathbf{x} = \mathbf{x}$, получаем: $(A - \lambda E)\mathbf{x} = A\mathbf{x} - \lambda\mathbf{x} = 0$. Отсюда следует, что число λ тогда и только тогда будет собственным значением линейного преобразования A, когда преобразование $A - \lambda E$ переводит некоторый вектор $\mathbf{x}$ пространства в вектор $\mathbf{0}$. Можно показать, что это возможно тогда и только тогда, когда $A - \lambda E$ является вырожденным линейным преобразованием, то есть когда определитель его матрицы равен 0.

Если рассматривать λ как неизвестное, то — в случае n-мерного пространства — упомянутый определитель будет многочленом n-ой степени от λ, который называется характеристическим многочленом данного линейного преобразования или соответствующей ему матрицы. Корни этого многочлена являются собственными значениями преобразования (или матрицы).

Если, например,

$[A] = \begin{pmatrix} -1 & 4 \\ 5 & -2 \end{pmatrix}$, то ее характеристический многочлен есть

$\begin{vmatrix} -1-\lambda & 4 \\ 5 & -2-\lambda \end{vmatrix} = (-1-\lambda)(-2-\lambda) - 20 = \lambda^2 + 3\lambda - 18$. Таким образом собственными значениями A являются числа 3 и -6.

Между матрицей и ее характеристическим многочленом существует очень тесная связь. Если характеристический многочлен преобразования A имеет вид

$$\lambda^n + \alpha_{n-1}\lambda^{n-1} + \ldots + \alpha_1\lambda + \alpha_0, \quad \text{то} \quad A^n + \alpha_{n-1}A^{n-1} + \ldots + \alpha_1 A + \alpha_0 E = 0,$$

то есть *всякое линейное преобразование является корнем своего характеристического многочлена*. В предыдущем примере

$$[A]^2 = \begin{pmatrix} (-1)(-1) + 4\cdot 5 & (-1)4 + 5(-2) \\ 5(-1) + (-2)5 & 5\cdot 4 + (-2)(-2) \end{pmatrix} = \begin{pmatrix} 21 & -12 \\ -15 & 24 \end{pmatrix},$$

$$3[A] = \begin{pmatrix} -3 & 12 \\ 15 & -6 \end{pmatrix} \quad \text{и} \quad -18[E] = \begin{pmatrix} -18 & 0 \\ 0 & -18 \end{pmatrix}, \quad \text{откуда}$$

$$[A^2 + 3A - 18E] = \begin{pmatrix} 21 - 3 - 18 & -12 + 12 + 0 \\ -15 + 15 + 0 & 24 - 6 - 18 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}.$$

СИСТЕМЫ УРАВНЕНИЙ

СИСТЕМЫ ЛИНЕЙНЫХ УРАВНЕНИЙ

Если в задаче требуется определить несколько неизвестных, между которыми существует несколько соотношений, то есть зависимость между которыми описывается несколькими уравнениями, то необходимо найти такие значения неизвестных, которые удовлетворяли бы каждому из уравнений. Однако в случае нескольких неизвестных решением является не одно число, а некоторая система чисел, в которой присутствуют некоторые конкретные значения каждого из неизвестных. Например, одним из решений уравнения $x^2 + xy - y^2 + yz + z + 1 = 0$ является совокупность значений $x = 1, y = -1, z = 2$ (или вектор с компонентами $(1; -1; 2)$).

Сначала мы займемся системами линейных уравнений. Системой линейных уравнений называется система уравнений вида

$$\begin{aligned} \alpha_{11}x_1 &+ \alpha_{12}x_2 + \ldots + \alpha_{1n}x_n = \beta_1 \\ \alpha_{21}x_1 &+ \alpha_{22}x_2 + \ldots + \alpha_{2n}x_n = \beta_2 \\ &\vdots \\ \alpha_{k1}x_1 &+ \alpha_{k2}x_2 + \ldots + \alpha_{kn}x_n = \beta_k. \end{aligned}$$

Решением этой системы уравнений называется всякая система чисел ξ_1, $\xi_2, \ldots, \xi_n$, при подстановке которых вместо соответствующих неизвестных каждое из уравнений обращается в тождество.

Системы линейных уравнений не всегда имеют решение. Если, например, в системе присутствуют (может быть «скрыто») уравнения $x = 1$ и $x = 2$, то очевидно, что такая система уравнений не имеет решения. Таким образом, необходимо отдельно исследовать вопрос, разрешима ли данная система. Однако, как правило, выяснение этого вопроса требует почти таких же усилий, как и само решение. Поэтому в тех случаях, когда необходимо найти решение, исследование разрешимости оказывается излишним. Мы изложим метод, с помощью которого такую систему можно решить, если она имеет решение (в этом случае система называется совместной), если же нет, то это выяснится в процессе решения.

Решение систем линейных уравнений. Если все коэффициенты в левых частях уравнений равны 0, то в зависимости от того, есть среди свободных членов (то есть чисел в правых частях) отличные от нуля или нет, система уравнений, соответственно, или не имеет решения, или же любая система n чисел является ее решением. Предположим теперь, что среди коэффициентов в левых частях уравнений есть отличные от 0. Поменяв, если нужно, некоторые из уравнений местами и перенумеровав неизвестные, можно добиться того, чтобы коэффициент α_{11} был отличен от нуля. Вычтем теперь первое уравнение, умноженное на $\frac{\alpha_{21}}{\alpha_{11}}$, из второго уравнения, затем первое уравнение, умноженное на $\frac{\alpha_{31}}{\alpha_{11}}$, вычтем из третьего уравнения и так далее, и, наконец, вычтем первое уравнение, умноженное на $\frac{\alpha_{k1}}{\alpha_{11}}$, из k-го уравнения (здесь имеется в виду вычитание каждой из частей одного уравнения из соответствующих частей другого). Очевидно, что все решения исходной системы уравнений будут решениями и для полученной системы. Верно и обратное утверждение, так как первое уравнение исходной системы совпадает с первым уравнением новой системы уравнений, а i-е уравнение ($i = 2, 3, \ldots, k$) исходной системы можно получить, если прибавить к i-му уравнению новой системы первое уравнение, умноженное на $\frac{\alpha_{i1}}{\alpha_{11}}$. В первом уравнении новой системы присутствуют такие неизвестные, которые не появляются ни в одном из остальных уравнений (таким неизвестным наверняка будет x_1). Предположим, что это неизвестные $x_1, x_2, \ldots, x_{r-1}$. Таким образом мы пришли к системе уравнений следующего вида:

$$\begin{aligned}
\alpha_{11}x_1 + \ldots + \alpha_{1r-1}x_{r-1} + \alpha_{1r}x_r + \ldots + \alpha_{1n}x_n &= \beta_1 \\
\gamma_{2r}x_r + \ldots + \gamma_{2n}x_n &= \delta_2 \\
&\vdots \\
\gamma_{kr}x_r + \ldots + \gamma_{kn}x_n &= \delta_k
\end{aligned}$$

Выясним, как можно получить решения этой системы уравнений. Если $(\xi_1, \ldots, \xi_{r-1}, \xi_r \ldots, \xi_n)$ является ее решением, то $(\xi_r, \ldots, \xi_n)$ будет решением системы уравнений, которая получится, если отбросить первое уравнение данной системы, и, следовательно, состоит из $k-1$ уравнений. Таким образом, сначала следует найти решения этой новой системы, состоящей из $k-1$ уравнений. Если $(\xi_1, \ldots, \xi_n)$ есть некоторое ее решение, то из первого уравнения исходной системы получается уравнение

$$\alpha_{11}x_1 + \ldots + \alpha_{1r-1}x_{r-1} = \beta_1 - \alpha_{1r}\xi_r - \ldots - \alpha_{1n}\xi_n.$$

Все решения этого уравнения можно получить, если, подставив вместо неизвестных $x_2, \ldots, x_{r-1}$ произвольные числа, решить уравнение

$$\alpha_{11}x_1 = \beta_1 - \alpha_{12}\xi_2 - \ldots - \alpha_{1n}\xi_n,$$

которое имеет единственное решение, поскольку $\alpha_{11} \neq 0$. Итак, всякая система линейных уравнений (в левых частях которых не все коэффициенты равны 0), может быть приведена к системе, содержащей на одно уравнение меньше. Продолжая этот процесс, мы получим, наконец, такую систему уравнений, где все коэффициенты в левых частях равны 0 (может, в частности, получится и система, «не содержащая ни одного уравнения»). Поскольку о таких системах сразу можно сказать, имеют ли они решения, и, если да, указать эти решения, то с помощью такого процесса можно найти решения любой системы линейных уравнений — если они существуют.

ПРИМЕР РЕШЕНИЯ СИСТЕМЫ ЛИНЕЙНЫХ УРАВНЕНИЙ

$$\begin{array}{r}
\underline{3x + y - 6t - u - v = 3} \\
2x + 2y - 4t + 3u + 6v = 5 \\
4x + y - 8t + u + v = 5 \\
-x + 2y + 2t + 2u + 2v = 1 \\
-5x + 4y + 10t + 6u + 12v = 1
\end{array}$$

Чтобы исключить неизвестное y, вычтем из второго уравнения первое, умноженное на 2, из третьего — первое, из четвертого — первое, умноженное на 2, а из пятого — первое уравнение, умноженное на 4 (первое уравнение останется без изменений, поэтому мы его подчеркнули):

$$\begin{array}{r}
-4x + 8t + 5u + 8v = -1 \\
\underline{x - 2t + 2u + 2v = 2} \\
-7x + 14t + 4u + 4v = -5 \\
-17x + 34t + 10u + 16v = -11.
\end{array}$$

Теперь с помощью второго уравнения исключим x (второе уравнение подчеркнуто, так как остается без изменений). Прибавляя к первому уравнению второе, умноженное на 4, к третьему — второе, умноженное на 7, а к четвертому — второе, умноженное на 17, получаем:

$$\begin{aligned} &+13u+16v=7 \\ &+18u+18v=9 \qquad \underline{+2u+2v=1} \\ &+44u+50v=23 \end{aligned}$$

Отсюда видно, что одновременно удалось исключить и t. Это означает, что, основываясь на данной системе уравнений, неизвестные x и t нельзя «разделить», удается лишь установить «связь» между ними. Чтобы упростить вычисления, разделим обе части второго уравнения на 9 и в дальнейшем оставим его без изменений. Следующим шагом можно исключить v, вычитая второе уравнение, умноженное на 8, из первого, а затем второе, умноженное на 25, из третьего:

$$\begin{aligned} -3u &= -1 \\ -6u &= -2. \end{aligned}$$

Если теперь вычесть из второго уравнения первое, умноженное на 2, то мы получим равенство $0=0$, а следовательно, система совместна. Одновременно это означает, что второе уравнение можно отбросить. Из первого уравнения определяем u: $\underline{u=\frac{1}{3}}$. Подставляя это значение в уравнение $2u+2v=1$, получаем: $\frac{2}{3}+2v=1$, откуда $\underline{v=\frac{1}{6}}$. Полученные значения u и v подставим в предыдущее подчеркнутое уравнение:

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

Отсюда приходим к соотношению $\underline{x-2t=1}$, то есть значения x и t действительно не удается найти. В то же время понятно, что последнему из рассмотренных уравнений удовлетворяют лишь такие пары значений x, t, для которых выполняется полученное соотношение. Отсюда следует также, что и в первом (подчеркнутом) уравнении неизвестные x и t не могут присутствовать отдельно, а лишь в выражении $x-2t$ с некоторым коэффициентом. Действительно, в первом уравнении стоит $3(x-2t)$. Подставляя полученные значения в первое уравнение, получаем:

$$3+y-\frac{1}{3}-\frac{1}{6}=3, \text{ то есть } \underline{y=\frac{1}{2}}.$$

Если теперь подставить вместо t некоторое конкретное значение t_0, то решение исходной системы уравнений можно записать в виде

$$x=1+2t_0, \quad y=\frac{1}{2}, \quad t=t_0, \quad u=\frac{1}{3}, \quad v=\frac{1}{6}.$$

В процессе решения мы убедились, что таким образом можно получить все решения, поэтому нет необходимости проверять подстановкой, удовлетворяет ли это решение исходной системе.

УСЛОВИЕ СУЩЕСТВОВАНИЯ РЕШЕНИЯ СИСТЕМЫ ЛИНЕЙНЫХ УРАВНЕНИЙ

нетрудно исследовать методами линейной алгебры. В самом деле, рассмотрим исходную систему уравнений. Выберем в k-мерном пространстве векторы $\mathbf{a}_1, \mathbf{a}_2, \ldots \mathbf{a}_n$ и $\mathbf{b}$ таким образом, чтобы компоненты вектора $\mathbf{a}_i$ совпадали с коэффициентами при x_i, а компоненты вектора $\mathbf{b}$ — со свободными членами уравнений (считая «сверху вниз»). Тогда решение системы уравнений равносильно отысканию таких чисел $\xi_1, \xi_2, \ldots, \xi_n$, для которых

$$\xi_1\mathbf{a}_1 + \xi_2\mathbf{a}_2 + \ldots + \xi_n\mathbf{a}_n = \mathbf{b}\,.$$

Таким образом система уравнений имеет решение тогда и только тогда, когда вектор $\mathbf{b}$ принадлежит к подпространству, порожденному векторами $\mathbf{a}_1, \mathbf{a}_2, \ldots \mathbf{a}_n$. Сформулируем это условие иначе. Рассмотрим матрицу из k строк и n столбцов, в которой j-ым элементом i-й строки является коэффициент при x_j в i-ом уравнении системы. Такую матрицу называют *матрицей системы уравнений*. Рассмотрим также матрицу из k строк и $n+1$ столбцов, получаемую приписыванием к матрице системы справа столбца свободных членов. Эта матрица называется *расширенной матрицей системы*. Справедливо следующее утверждение:

система линейных уравнений тогда и только тогда имеет решения, когда ранг ее матрицы равен рангу расширенной матрицы.

Понятие ранга матрицы упоминалось при изучении линейной алгебры. Там было, в частности, отмечено, что ранг матрицы не изменится, если к одной из ее строк (столбцов) прибавить другую строку (столбец), умноженную на некоторое число. Такими преобразованиями можно привести матрицу к виду, где в любой строке и в любом столбце имеется не более одного отличного от нуля элемента. Поскольку при этих преобразованиях ранг матрицы не меняется, то достаточно определить ранг преобразованной матрицы. Нетрудно показать, что если в каждой строке и в каждом столбце матрицы имеется не более одного отличного от нуля элемента, то ранг матрицы равен числу ее элементов, отличных от нуля. При решении системы линейных уравнений может встретиться случай, когда значения некоторых неизвестных выбираются произвольным образом. Если имеются такие неизвестные, то система уравнений имеет больше одного решения. Часто возникает вопрос, имеет ли данная система единственное решение. Оказывается, что

система линейных уравнений тогда и только тогда имеет единственное решение, когда она совместна, и ранг ее матрицы равен числу неизвестных.

В частном случае, когда число уравнений равно числу неизвестных, то есть, когда матрица системы является квадратной, *для того, чтобы совместная система линейных уравнений имела единственное решение, необходимо и достаточно, чтобы определитель ее матрицы был отличен от нуля.*

Правило Крамера. В последнем случае систему линейных уравнений можно решить и с помощью определителей. Решение находится по так называемому *правилу Крамера:*

$$x_j = \frac{D_j}{D}, \qquad j = 1, 2, \ldots, n,$$

где D — определитель матрицы системы, а D_j — определитель, который получается из D заменой в нем j-го столбца столбцом свободных членов. Однако этот способ решения требует бóльших вычислений, чем метод, изложенный выше.

СИСТЕМЫ ЛИНЕЙНЫХ ОДНОРОДНЫХ УРАВНЕНИЙ

Самостоятельный интерес представляют системы линейных уравнений, в которых все свободные члены равны 0. Такие системы называются системами однородных линейных уравнений.

В случае систем однородных линейных уравнений вопрос о совместности системы отпадает, так как такие системы всегда имеют так называемое *тривиальное решение,* в котором значения всех неизвестных равны нулю. В то же время здесь возникает вопрос, когда система имеет нетривиальное решение. Ответ на этот вопрос следующий:

система однородных линейных уравнений тогда и только тогда имеет нетривиальное решение, когда ранг ее матрицы меньше числа неизвестных.

Если число уравнений системы однородных линейных уравнений равно числу неизвестных, то необходимым и достаточным условием существования нетривиального решения является равенство нулю определителя матрицы системы.

Оба эти результата немедленно вытекают из условия существования единственного решения системы линейных уравнений (точнее из обратного условия, когда система имеет больше одного решения). Чаще всего (особенно в геометрии при исследовании уравнений геометрических фигур) последний результат используется как раз в обратнйо форме:

если система однородных линейных уравнений, число уравнений в которой равно числу неизвестных, имеет нетривиальное решение, то определитель ее матрицы равен нулю.

НЕСОВМЕСТНЫЕ* СИСТЕМЫ УРАВНЕНИЙ

На практике часто встречаются задачи, в которых неизвестные величины требуется определить путем измерений. Однако, чтобы устранить влияние ошибок измерений, производят больше измерений, чем это необходимо

* В оригинале используется название «переопределенные» системы.

для определения неизвестных. В результате мы приходим к так называемой *несовместной системе уравнений*, которая не имеет решений. Мы исследуем случай, когда получаемая система оказывается линейной. Если перейти к векторным обозначениям, то исходная задача состоит в следующем: требуется найти такие значения $\xi_1, \xi_2, \ldots, \xi_n$, для которых $\xi_1\mathbf{a}_1 + \xi_2\mathbf{a}_2 + \ldots + \xi_n\mathbf{a}_n = \mathbf{b}$. Однако, как мы отметили, таких значений не существует. По этой причине вместо них разыскивают такие значения $\eta_1, \eta_2, \ldots, \eta_n$, для которых вектор $\mathbf{a} = \eta_1\mathbf{a}_1 + \eta_2\mathbf{a}_2 + \ldots \eta_n\mathbf{a}_n$ как можно меньше отличается от вектора $\mathbf{b}$. Таким свойством обладает «ортогональная проекция» вектора $\mathbf{b}$ на подпространство, порожденное векторами $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n$. Другими словами, вектор $\mathbf{a}$ следует определить таким образом, чтобы вектор $\mathbf{a} - \mathbf{b}$ был ортогонален каждому из векторов $\mathbf{a}_i$, то есть должны выполняться равенства $(\mathbf{a} - \mathbf{b}, \mathbf{a}_i)$ или $(\mathbf{a}, \mathbf{a}_i) = (\mathbf{b}, \mathbf{a}_i)$. Записывая вместо вектора $\mathbf{a}$ его выражение через векторы $\mathbf{a}_i$, приходим к уравнениям вида

$$\eta_1(\mathbf{a}_1, \mathbf{a}_i) + \eta_2(\mathbf{a}_2, \mathbf{a}_i) + \ldots + \eta_n(\mathbf{a}_n, \mathbf{a}_i) = (\mathbf{b}, \mathbf{a}_i) \qquad (i = 1, \ldots, n).$$

Можно показать, что эта система уравнений всегда имеет решение, причем если векторы $\mathbf{a}_i$ линейно независимы (что обычно выполняется), то решение оказывается единственным.

Проиллюстрируем сказанное на примере. Пусть имеется следующая (несовместная) система уравнений:

$$\begin{aligned} 2x + 3y &= 5 \\ 3x + y &= 5, \\ 2x + 5y &= 6 \end{aligned} \quad \text{то есть} \quad [\mathbf{a}_1] = \begin{bmatrix} 2 \\ 3 \\ 2 \end{bmatrix} \quad [\mathbf{a}_2] = \begin{bmatrix} 3 \\ 1 \\ 5 \end{bmatrix} \quad [\mathbf{b}] = \begin{bmatrix} 5 \\ 5 \\ 6 \end{bmatrix}.$$

Вычислим соответствующие скалярные произведения: $(\mathbf{a}_1, \mathbf{a}_1) = 4 + 9 + 4 = 17$, $(\mathbf{a}_2, \mathbf{a}_2) = 9 + 1 + 25 = 35$, $(\mathbf{a}_1, \mathbf{a}_2) = (\mathbf{a}_2, \mathbf{a}_1) = 6 + 3 + 10 = 19$, $(\mathbf{b}, \mathbf{a}_1) = 10 + 15 + 12 = 37$ и $(\mathbf{b}, \mathbf{a}_2) = 15 + 5 + 30 = 50$. Таким образом, для искомых значений мы получаем следующую систему уравнений:

$$17x + 19y = 37$$

$$19x + 35y = 50,$$

решение которой есть $x = \dfrac{115}{78}$, $y = \dfrac{49}{78}$. Подставляя эти значения в уравнения исходной системы, мы получим, соответственно, $\dfrac{377}{78}$, $\dfrac{394}{78}$ и $\dfrac{485}{78}$.

Величину отклонения можно определить и из геометрических соображений. Рассмотрим вектор, компонентами которого являются полученные величины: $\left(\dfrac{377}{78}, \dfrac{394}{78}, \dfrac{485}{78}\right)$, а также вектор с компонентами, равными, соответственно, свободным членам исходных уравнений: (5, 5, 6).

Расстояние между концами этих векторов, равное квадратному корню из суммы квадратов разностей соответственных компонентов, оказывается немного меньше 0,28. Таким образом погрешность решения равна 0,28.

СИСТЕМЫ УРАВНЕНИЙ ВЫСШИХ СТЕПЕНЕЙ

Итак, можно считать, что решение систем линейных уравнений в принципе осуществимо. В действительности, однако, положение несколько иное, так как системы уравнений, встречающиеся на практике, нередко содержат несколько сотен неизвестных. В таких случаях решение систем уравнений практически неосуществимо. Большую помощь при решении систем линейных уравнений оказывают электронные вычислительные машины, однако, эту проблему не всегда удается решить даже с их помощью.

Значительно хуже обстоит дело с решением систем уравнений высших степеней. Для таких систем неизвестен даже метод, с помощью которого можно было бы установить какое-либо необходимое и достаточное условие их разрешимости. Однако можно указать такой алгоритм, с помощью которого систему можно решить, если она вообще разрешима, если же нет, то это выясняется в процессе решения. К сожалению, этот алгоритм имеет лишь принципиальное значение, так как в случае нескольких неизвестных система уравнений уже через несколько шагов становится очень сложной.

Суть этого алгоритма в следующем: если имеется k уравнений с n неизвестными, то эта система приводится к системе $k-1$ уравнений с $n-1$ неизвестными. Однако это дается дорогой ценой, так как при этом степени промежуточных уравнений «почти перемножаются», что и делает такой метод решения чрезвычайно сложным.

Прежде чем перейти к изложению этого метода, напомним несколько теорем о многочленах. Если $f(x)$ и $g(x)$ — многочлены, соответственно, n-й и k-й степени, имеющие общие корни, то найдутся такие (не нулевые) многочлены, $u(x)$ и $v(x)$, соответственно, $(k-1)$-й и $(n-1)$-й степени, что $f(x)\,u(x) - g(x)\,v(x) = 0$. Отсюда для коэффициентов многочленов $u(x)$ и $v(x)$ получается система однородных линейных уравнений, которая, согласно сказанному, имеет ненулевое решение. Следовательно, определитель матрицы системы (являющейся квадратной матрицей) равен 0. Если при этом коэффициенты многочленов $f(x)$ и $g(x)$ сами являются многочленами от какого-то неизвестного, то из равенства нулю определителя мы получим уравнение относительно этого неизвестного. Подставляя его решения в исходное уравнение, можно найти общие корни многочленов $f(x)$ и $g(x)$, соответствующие каждому решению.

Объясним без доказательства, какой вид будет иметь упомянутая матрица. В первых k ее строках будут стоять только коэффициенты многочлена $f(x)$. В первой строке сначала по порядку расположены коэффициенты начиная со старшего и кончая свободным членом, после чего стоит столько нулей, сколько необходимо, чтобы число столбцов матрицы равнялось $n+k$. Во второй строке на первом месте стоит 0, затем — в том же порядке — коэффициенты многочлена $f(x)$, а на остальных местах нули. В каждой из последующих строк также располагаются коэффициенты $f(x)$, но всякий раз «со сдвигом на одно место вправо», а остальные места заполняются нулями. В k-й строке уже все нули «перейдут» в начало. Остальные n строк аналогичным образом заполняются коэффициентами многочлена $g(x)$ и нулями.

Пример: пусть имеется следующая система:

$$(1)\ x^2 + 2x - y^2 + 4y - 3 = 0 \qquad (2)\ x^2 + 3xy + 2y^2 + y - 1 = 0.$$

Рассматривая левые части этих уравнений как многочлены от x, получаем следующий определитель:

$$\begin{vmatrix} 1 & 2 & -y^2 + 4y - 3 & 0 \\ 0 & 1 & 2 & -y^2 + 4y - 3 \\ 1 & 3y & 2y^2 + y - 1 & 0 \\ 0 & 1 & 3y & 2y^2 + y - 1 \end{vmatrix}.$$

Этот определитель должен быть равен 0. Вычисляя его, приходим к уравнению $12y^3 - 28y^2 + 16y = 0$. Вынося y за скобку и решая полученное квадратное уравнение, получаем следующие корни: $y = 0$, $y = 1$ и $y = \frac{4}{3}$. Подставляя эти значения в исходную систему, получаем, соответственно, следующие уравнения относительно x:

$$x^2 + 2x - 3 = x^2 - 1 = 0; \qquad x^2 + 2x = x^2 + 3x + 2 = 0;$$

$$x^2 + 2x + \frac{5}{9} = x^2 + 4x + \frac{35}{9} = 0.$$

Их общими решениями являются, соответственно, $x = 1$, $x = -2$ и $x = -\frac{5}{3}$. В том, что это действительно так, можно убедиться прямой подстановкой. Таким образом решениями исходной системы уравнений являются следующие пары чисел: $(1, 0)$, $(-2, 1)$, $\left(\frac{4}{3}, -\frac{5}{3}\right)$.

ТЕОРИЯ ГРУПП

Теория групп является одной из старейших ветвей алгебры. Кроме самой алгебры, в качестве областей ее применения можно назвать почти любой раздел математики. Мы, естественно, ограничимся здесь лишь знакомством с элементами теории групп; ее применений, учитывая их многообразие и сложность, в силу необходимой сжатости изложения мы не будем касаться.

ПОНЯТИЕ ГРУППЫ

Группа — точно так же, как векторное пространство — является алгебраической структурой (алгебраической системой). Наглядно группу можно было бы охарактеризовать следующим образом: если в некоторой алгебраической структуре рассматривается только одна операция, например, сложение (или умножение), причем вычитание (деление) всегда выполнимо, то можно говорить о группе. Такими алгебраическими структурами являются, например, множество целых чисел относительно сложения или множество положительных чисел относительно умножения. Такую структуру образуют и преобразования плоскости, сохраняющие расстояние, если в качестве операции рассмотреть последовательное выполнение преобразований.

Аксиомы группы. Естественно, что и в случае групп не следует привязываться к конкретным свойствам отдельных примеров. По этой причине необходимо задать свойства операции, определенной в группе, при помощи аксиом.

Некоторое (непустое) множество G называется группой, если выполняются следующие четыре аксиомы:

(1) в множестве G определена некоторая операция над его элементами, которую мы назовем умножением, то есть любым двум элементам a и b из G поставлен в соответствие некоторый третий элемент из G, который обозначается через ab и называется произведением элементов a и b;

(2) умножение ассоциативно, то есть для любых трех элементов a, b и c из G $(ab)c = a(bc)$;

(3) в G существует единичный элемент, то есть такой элемент e, что для любого a из G $ea = ae = a$;

(4) для любого элемента в группе существует обратный ему элемент, то есть для любого элемента a найдется такой элемент a^{-1}, для которого $(a^{-1})\ a = a(a^{-1}) = e$ (где e определен выше).

Сразу заметим, что при умножении следует учитывать порядок элементов, так как, вообще говоря, ab и ba могут быть различными.

Если для любых двух элементов a и b группы выполнено равенство $ab = ba$, то группа называется коммутативной или абелевой.

Естественно, может случиться, что в некоторой группе традиционно обозначают операцию иначе, например, $a + b$ или $a \cdot b$. Однако это не имеет принципиального значения, важно лишь, обладает ли эта операция указанными выше свойствами.

Нетрудно доказать следующие утверждения: никакой другой элемент, кроме e, не обладает свойством, указанным в аксиоме (3); для всякого элемента a существует единственный обратный ему элемент; для любых двух элементов a и b группы найдется единственный элемент g и единственный элемент h, для которых $ag = b$ и $ha = b$.

ГРУППЫ ПОДСТАНОВОК

Теория групп возникла при исследовании разрешимости алгебраических уравнений. Для этого появилась необходимость рассмотреть некоторые перестановки корней уравнения, и выяснилось, что эти перестановки относительно некоторой операции образуют группу. По этой причине мы рассмотрим этот пример несколько подробнее.

С перестановками мы уже встречались при ознакомлении с комбинаторикой. Под перестановкой понималось расположение данных элементов в некотором порядке. Можно предположить, что элементы, о которых идет речь, являются числами $1, 2, \ldots, n$. Записывая эти числа в какомлибо порядке, мы получаем некоторую их перестановку. Исследуем, однако, сам процесс, с помощью которого из исходной перестановки мы получаем новую. Рассмотрим, например, перестановку 3, 4, 7, 2, 1, 5, 6 чисел 1, 2, 3, 4, 5, 6, 7. Процесс образования этой перестановки можно описать следующим образом: запишем не первое место число 3, на второе — 4, на третье — 7 и так далее. Однако целесообразнее указывать не на то, какое из чисел записывается на какое место, а на то, *какое из чисел попадает на место какого числа*. Так, на место числа 1 попадает 3, на место числа 2 — число 4, на место 3 — число 7 и так далее. Для обозначения этого соответствия условимся под каждым из чисел исходной перестановки записывать число, которое попадает на его место в новой перестановке. При этом мы получим некоторую матрицу, которая в рассмотренном примере имеет следующий вид: $\begin{bmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ 3 & 4 & 7 & 2 & 1 & 5 & 6 \end{bmatrix}$. Итак, эта матрица обозначает переход от одной перестановки к другой. Такой переход называется подстановкой. Заметим, что здесь уже не существенно, какое из чисел стоит на каком месте. Важно лишь, какое из чисел стоит *под каким*, поэтому можно записать в произвольном порядке и числа верхней строки, но тогда надо соответствующим образом переставить и числа в нижней строке, чтобы пары остались неизменными. Таким образом ту же самую подстановку обозначает, например, и запись $\begin{bmatrix} 3 & 5 & 2 & 1 & 7 & 6 & 4 \\ 7 & 1 & 4 & 3 & 6 & 5 & 2 \end{bmatrix}$.

Умножение подстановок. Определим теперь умножение двух подстановок как их последовательное выполнение. Обозначим рассмотренную выше подстановку через π и рассмотрим подстановку $\varrho = \begin{bmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ 3 & 5 & 2 & 1 & 7 & 6 & 4 \end{bmatrix}$.

По определению произведение $\varrho\pi$ получается, если сначала выполнить подстановку ϱ, а затем — подстановку π. Посмотрим, что попадает при этом на место каждого из чисел 1, 2, 3, 4, 5, 6, 7. Подстановка ϱ ставит на место числа 1 число 3. Однако нужно еще выполнить подстановку π. Эта подстановка для каждого из чисел указывает, какое число должно попасть на его место, в частности, на место числа 3 попадает 7. В результате, если последовательно выполнить обе подстановки, то на место числа 1 попадет число 7 (говорят, что число 1 «переходит» в число 7). Рассмотрим теперь число 2. Подстановка ϱ переводит 2 в 5. После этого следует выполнить подстановку π, переводящую 5 в 1. В результате 2 переходит в 1. Смотрим дальше: 3 переходит в 2, затем 2 переходит в 4 — в результате 3 переходит в 4. 4 переходит в 1, затем 1 переходит в 3 — в результате 4 переходит в 3. 5 переходит в 7, а 7 переходит в 6 — в результате 5 переходит в 6. 6 переходит в 6, затем 6 переходит в 5 — в результате 6 переходит в 5. 7 переходит в 4, затем 4 переходит в 2 — в результате 7 переходит в 2. Таким образом имеем:

$$\begin{bmatrix}1&2&3&4&5&6&7\\3&5&2&1&7&6&4\end{bmatrix}\cdot\begin{bmatrix}3&5&2&1&7&6&4\\7&1&4&3&6&5&2\end{bmatrix}=\begin{bmatrix}1&2&3&4&5&6&7\\7&1&4&3&6&5&2\end{bmatrix}$$

Нетрудно видеть, что в результате мы получили некоторую подстановку. Очевидно, это справедливо и в общем случае. Больше того, можно сформулировать и общее правило получения произведения двух подстановок. Пусть мы хотим перемножить две подстановки. Запишем вторую подстановку таким образом, чтобы верхняя строка ее записи совпадала с нижней строкой записи первой подстановки. Тогда произведение запишется в виде двух строк, из которых верхняя совпадает с верхней строкой записи первой подстановки, а нижняя — с нижней сторокой записи второй подстановки. (Заметим, что мы нигде не используем того факта, что верхняя строка первой подстановки записана в порядке возрастания чисел.) Сразу следует отметить, что такое умножение подстановок некоммутативно. Рассмотрим, например, произведение $\pi\varrho$:

$$\begin{bmatrix}1&2&3&4&5&6&7\\3&4&7&2&1&5&6\end{bmatrix}\cdot\begin{bmatrix}1&2&3&4&5&6&7\\3&5&2&1&7&6&4\end{bmatrix}=\begin{bmatrix}1&2&3&4&5&6&7\\2&1&4&5&3&7&6\end{bmatrix}.$$

Покажем, что относительно такого умножения подстановки образуют группу. Аксиома (1) выполнена, так как умножение определено. Рассмотрим теперь числа $1, 2, 3, \ldots, n$ и пусть P — некоторая их подстановка. Рассмотрим некоторое число i (и здесь, и в дальнейшем рассматриваются только числа $1, 2, \ldots, n$). То число, которое при подстановке P переходит в i, обозначим через $(i)P$. Пусть теперь P и Q — две произвольные подстановки. Посмотрим, какое число переходит в i при подстановке PQ. По определению умножения сначала следует произвести подстановку P, а затем — Q. При подстановке P на место i попадает число $(i)P$. При подстановке Q на место произвольного числа j попадает число $(j)Q$ и, в частности, на место числа $(i)P$ попадает число $[(i)P]Q$. Таким образом $(i)(PQ) = [(i)P]Q$. Теперь можно доказать ассоциативность умножения, то есть выполнение аксиомы (2). Нужно показать, что для любых подстановок P, Q и R выполняется равенство $(PQ)R = P(QR)$ или, другими словами, что при подстановках $(PQ)R$ и $P(QR)$ на место числа i попадает одно и то же число. Согласно установленному выше,

$$(i)[(PQ)R] = [(i)(PQ)]R = [((i)P)Q]R,$$

$$(i)[P(QR)] = [(i)P](QR) = ([(i)P]Q)R.$$

Полученные числа действительно равны. Теперь нужно показать существование единицы, то есть выполнение аксиомы (3). Мы утверждаем, что единицей является так называемая *тождественная подстановка*, то есть такая подстановка E, при которой $(i)E = i$. Это действительно подстановка; в ее записи под каждым из чисел стоит оно само, то есть при подстановке E все числа остаются на своих местах. Пусть P — произвольная подстановка, тогда $(i)(EP) = [(i)E]P = (i)P$ и $(i)(PE) = [(i)P]E = (i)P$. Таким образом $EP = PE = P$. Докажем, наконец, выполнение аксиомы (4). Пусть P — произвольная подстановка. Обратную ей подстановку определим следующим образом: если P переводит i в j, то пусть обратная ей подстановка переводит j в i. Таким образом запись обратной подстановки получается из записи P переменой строк. Поскольку в записи P в каждой из строк присутствуют все числа, то то же справедливо и для записи обратной подстановки. Итак, p^{-1} есть такая подстановка, что $[(i)P]P^{-1} = i$. Отсюда сразу следует, что $(i)(PP^{-1}) = [(i)P]P^{-1} = i = (i)E$, то есть $PP^{-1} = E$. Рассмотрим теперь $(i)(P^{-1}P)$. Поскольку P — некоторая подстановка, то существует некоторое число j, для которого $(j)P = i$. Отсюда:

$$(i)(P^{-1}P) = ([j]P)(P^{-1}P) = (j)(PP^{-1}P) = (j)(EP) = (j)P = i,$$

то есть $P^{-1}P = E$.

ЦИКЛЫ

Группа всех подстановок из n элементов называется *симметрической группой* и обозначается через S_n.

Число элементов группы S_n совпадает с числом всех перестановок из n элементов, то есть равно $n!$ При записи подстановок в виде двух строк каждое число присутствовало дважды. Однако этого можно избежать, если составить одну строку, записывая после каждого из чисел то число, в которое оно переходит. Так, например, при рассмотренной выше подстановке π число 1 переходит в 3, поэтому запишем: (13. В свою очередь число 3 переходит в 7, которое записываем после 3: (137. Число 7 переходит в 6, 6 — в 5, поэтому имеем: (13765. Число 5 переходит в 1, которое таким образом появляется вторично. Записывать его мы не будем, а закроем скобку: (13765). Этим мы еще не определили для каждого числа, в какое число оно переходит. Однако и это нетрудно сделать. Рассмотрим первое из отсутствующих чисел: это число 2, которое переходит в 4, в то же время 4 переходит в 2, что обозначается следующим образом: (24). В результате всю подстановку можно задать в виде (13765)(24).

Такой способ записи подстановок называется *разложением в циклы*, при этом каждая группа чисел, стоящая в скобках, называется *циклом*.

Разложение в циклы позволяет глубже раскрыть механизм подстановки. Нетрудно видеть, например, что указанную выше подстановку можно по сути дела рассматривать как две подстановки. Одна из них является подстановкой чисел 1, 3, 7, 6, 5, для выполнения которой нужно на место каждого из чисел поставить следующее за ним число, а на место последнего числа — первое. Если представить себе эти числа расположенными на окружности через равные промежутки, то выполнение подстановки равносильно повороту окружности «на один промежуток». Вторая подстановка означает перемену местами чисел 2 и 4. Заметим, что при записи цикла не существенно, какое из чисел стоит на первом месте, важен только их порядок, таким образом (13765) = (37651) = (76513) = = (65137) = (51376). При рассмотрении исходной подстановки можно было сначала посмотреть, в какое число переходит 2, тогда мы получили бы запись (24) (13765). Таким образом для этих циклов имеет место своего рода «коммутативность», и это понятно, так как они переставляют разные числа и, следовательно, «не мешают друг другу». Подстановка ϱ разлагается в циклы следующим образом: (132574)(6). Цикл, состоящий из одного элемента 6, можно отбросить, если трактовать запись таким образом, что отсутствующие числа остаются на месте. (Так, например, можно считать, что при рассмотренных подстановках остаются на месте числа 8, 9, . . .). Такая запись позволяет выполнять и умножение подстановок. Например, произведение $\pi\varrho$ можно рассматривать как произведение трех сомножителей: (13765) (24) (132574). Число 1 сначала переходит в 3, второй сомножитель оставляет 3 на месте, а третий переводит 3 в 2. Коротко это можно обозначить так: 1 — 3 — 3 — 2. Теперь рассмотрим число 2: 2 — 2 — 4 — 1. Тем самым получен «первый» цикл произведения: (12). Первое из оставшихся чисел есть число 3; для него имеем: 3 — 7 — 7 — — 4, 4 — 4 — 2 — 5, 5 — 1 — 1 — 3. В результате получаем следующий цикл: (345). Наконец, 6 — 5 — 5 — 7, 7 — 6 — 6 — 6. Таким образом все произведение имеет вид (12)(345)(67). Отсюда легко восстановить запись из двух строк, в результате мы получим матрицу, уже рассмотренную выше. Произведение $\varrho\pi$ получается следующим образом: (132574) (13765) (24) = (172) (34) (56).

Транспозиция. Из рассмотренного процесса очевидно, что *всякая подстановка может быть разложена в произведение циклов. Число элементов цикла называется его длиной. Цикл длины 2 называется транспозицией.* Очевидно, что транспозиция означает, что два элемента меняются местами. Транспозиции являются как бы «кирпичиками», из которых строятся подстановки. *Всякая подстановка может быть разложена в произведение транспозиций.* Очевидно, что достаточно доказать это утверждение для циклов. Рассмотрим, например, цикл (13765). Здесь 1 переходит в 3, этого можно достигнуть транспозицией (13). На следующем шагу 3 переходит в 7. Однако при выполнении предыдущей транспозиции числа 1 и 3 поменялись местами, поэтому 7 должно теперь попасть на место 1, что достигается транспозицией (17). Путем аналогичных рассуждений приходим к следующему разложению: (13765) = (13) (17) (16) (15). Представим себе опять, что эти числа расположены на некоторой окружности. Тогда полученный результат можно интерпретировать следующим образом: число 1 по очереди меняется местами с каждым из чисел, в результате чего окружность поворачивается «на одну пятую». Отсюда следует, что любая перестановка чисел может быть получена из перестановки, в которой числа расположены в порядке возрастания, путем последовательного выполнения некоторых транспозиций. Можно показать, что если при этом совершается четное число транспозиций, то получается четная, а при нечетном числе транспозиций — нечетная перестановка. Нетрудно показать также, что четные подстановки (то есть подстановки, при которых из чисел, расположенных в порядке возрастания, образуется четная перестановка) образуют группу относительно введенного выше умножения. Эта группа называется *знакопеременной* или *альтернирующей группой подстановок из n элементов* и обозначается через A_n.

СООТВЕТСТВИЯ МЕЖДУ ГРУППАМИ

Понятие группы мы определили с помощью аксиом. При этом не принимались во внимание ни природа элементов группы, ни сущность определенной в ней операции. По этой причине мы не можем отличить друг от друга две группы, если между их элементами удается установить такое соответствие, при котором произведение соответственных элементов были бы элементами, соответствующими друг другу. Такие группы называются изоморфными. В алгебре можно исследовать только такие свойства, которые не меняются при замене группы некоторой другой группой, изоморфной с ней.

Дадим точное определение изоморфизма:
группы G и G' называются изоморфными, если между ними можно установить взаимно однозначное соответствие, сохраняющее операцию.

Это означает, что каждому элементу a группы G можно однозначно поставить в соответствие некоторый элемент a' группы G' таким образом, что для любого элемента b' группы G' найдется ровно один такой элемент b группы G, которому соответствует как раз элемент b'; кроме того, если элементу a соответствует a', а элементу b соответствует b', то произведению ab соответствует произведение $a'b'$. Пусть, например, $G = S_2$. Эта группа состоит из двух элементов: тождественной подстановки E и подстановки $P = (12)$. Для них выполнены следующие свойства: $EE = E, EP = PE = P, PP = E$. Выберем в качестве G', группу, состоящую из чисел $+1$ и -1 (относительно обычного умножения чисел). Имеем: $(+1)(+1) = +1, (+1)(-1) = (-1)(+1) = -1$ и $(-1)(-1) = +1$. Таким образом соответствие $E \leftrightarrow +1$, $P \leftrightarrow -1$ является изоморфизмом.

Если группы G и G', изоморфны, то это обозначается следующим образом: $G \cong G'$.

Подмножество H группы G называется подгруппой группы G, если H само является группой относительно операции, определенной в G.

Например, упомянутая выше группа A_n является подгруппой группы S_n. Комплексные числа образуют группу относительно сложения; ее подгруппы (относительно сложения) образуют вещественные числа; подгруппой группы вещественных чисел является группа рациональных чисел; подгруппу последней группы образуют целые числа.

Гомоморфизм. В теории групп очень важную роль играет понятие гомоморфизма, то есть отображения, при котором, в отличие от изоморфизма, предполагается однозначность только в одну сторону.

Группа G' называется гомоморфным образом группы G, если существует однозначное отображение группы G в группу G', сохраняющее операцию. Такое отображение называется гомоморфизмом и обозначается следующим образом: $G \simeq G'$.

Итак, при гомоморфизме каждому элементу a группы G ставится в соответствие некоторый элемент a' группы G' таким образом, что для всякого элемента b' группы G' найдется некоторый элемент b группы G, которому он соответствует (но в G уже может быть несколько элементов, которым соответствует b'!); кроме того, если элементу a соответствует a', а элементу b — b', то элементу ab соответствует $a'b'$. Например, гомоморфным образом группы S_n является группа, состоящая из элементов $+1$ и -1 (относительно умножения), если четным подстановкам поставить в соответствие число $+1$, а нечетным — число -1. Другой пример: гомоморфным образом группы комплексных чисел относительно сложения является группа вещественных чисел (если комплексному числу $a + bi$ поставить в соответствие вещественное число a).

Для любой группы всегда можно указать два ее гомоморфных образа. Один из них — сама группа; здесь при гомоморфизме группы на себя каждому элементу ставится в соответствие он сам. Другим является группа, состоящая из одного элемента; в этом случае всем элементам группы ставится в соответствие этот элемент (единственным элементом группы, состоящей из одного элемента, является, очевидно, единичный элемент; эта группа изоморфна группе, состоящей из элемента $(+1)$). Эти два гомоморфизма называются *тривиальными гомоморфизмами*. В отличие от них все остальные гомоморфизмы называются *нетривиальными гомоморфизмами*.

Если для группы не существует нетривиального гомоморфизма, то группа называется примарной.

Примарной является всякая группа, число элементов которой есть простое число (например, группы, состоящие из 2, 3, 5 и так далее элементов). Можно показать, что если $n \neq 4$, то группа A_n является примарной; если же $n \geqq 5$, то A_n является примарной и некоммутативной группой.

Значение групп подстановок показывает следующий результат:

теорема Кэли: любая группа из n элементов изоморфна некоторой подгруппе группы S_n.

ТЕОРИЯ ПОЛЯ

ПОНЯТИЕ ПОЛЯ

Мы уже встречались с числовыми полями. Аналогично векторным пространствам и группам, поля также определяются с помощью аксиом. *Аксиомы поля.* Множество K называется полем, если оно состоит по крайней мере из двух элементов и если для его элементов

(1) определена операция сложения,	(1') определена операция умножения,
(2) сложение ассоциативно,	(2') умножение ассоциативно,
(3) сложение коммутативно,	(3') умножение коммутативно,
(4) выполнимо вычитание,	(4') выполнимо деление (кроме деления на 0).

Таким образом, для любых двух элементов a и b поля найдется такой

$$\text{элемент } x, \text{ что } a + x = b; \qquad \text{элемент } y, \text{ что } ay = b, \text{ если } a \neq 0.$$

В поле выполняется дистрибутивность, то есть $(a + b)c = ac + bc$.

Очевидно, что в любом числовом поле выполнены все перечисленные аксиомы.

Поле L называется расширением поля K, если L содержит все элементы из K. Так, расширением поля рациональных чисел является поле вещественных чисел; расширением поля вещественных чисел является поле комплексных чисел.

Элемент α поля L называется алгебраическим над полем K, если существует такой ненулевой многочлен с коэффициентами из K, для которого α является корнем. Если такого многочлена не существует, то элемент α называется трансцендентным над полем K.

Можно показать, что для любого ненулевого многочлена с коэффициентами из поля K существует такое расширение поля K, в котором этот многочлен имеет корень. Больше того, можно доказать, что для любого поля K существует такое его расширение, в котором любой многочлен с коэффициентами из K имеет корень. В современной алгебре эта теорема заменяет основную теорему алгебры, так как обеспечивает существование такого расширения поля комплексных чисел, которое обладает необходимыми алгебраическими свойствами.

ТЕОРИЯ ГАЛУА

С помощью теории поля открывается возможность для исследования разрешимости уравнений. Однако прежде всего необходимо выяснить, что понимается под решением некоторого уравнения. Возможны три случая:

1. О корнях уравнения мы хотим знать ровно столько, сколько необходимо чтобы производить вычисления с ними (это используется, например, при «исключении корней из знаменателя»).
2. Для корней уравнения мы хотели бы получить формулы, аналогичные известным формулам для корней уравнений второй, третьей и четвертой степени.
3. Корни уравнения мы хотим определить с заранее заданной точностью (это особенно важно для практических задач, когда, с одной стороны, нет необходимости в точном значении, с другой стороны, точного значения и нельзя получить из-за ошибок данных).

Случай 1 исчерпывается применением расширений полей. Согласно этому, о корнях уравнения предполагается лишь, что они «удовлетворяют данному уравнению». Например, при вычислениях с $\sqrt{2}$ можно использовать лишь тот факт, что $(\sqrt{2})^2 - 2 = 0$. В случае 3 решение можно найти с помощью различных численных методов, однако это уже не относится к алгебре. Во втором случае исчерпывающий ответ дается с помощью *тео-*

рии Галуа. Эту теорию создал французский математик Эварист Галуа. Свои результаты он записал в возрасте 21 года в ночь перед дуэлью, на которой погиб. Суть теории в том, что исследуются определенные перестановки корней, образующие группу. Такая группа получила название *группы Галуа*, соответствующей данному уравнению. Из строения этой группы можно с учетом знаков корней сделать вывод о разрешимости уравнения. Как мы видели, для уравнений не выше четвертой степени существует формула решения. Оказывается, что для уравнений выше четвертой степени таких формул не существует. Это является следствием результата, устанавливаемого в теории групп: при $n \geqq 5$ группа A_n является примарной. Сам по себе этот факт имеет лишь принципиальное значение, так как уже в случае уравнений четвертой степени формулы корней практически не применимы.

ВОЗМОЖНОСТЬ ГЕОМЕТРИЧЕСКИХ ПОСТРОЕНИЙ

Благодаря теории Галуа появляется возможность исследовать разрешимость некоторых задач на построение. Под построением понимается некоторый процесс, при котором из данных точек с помощью заданных средств за конечное число шагов находятся новые точки. Эта задача также не представляет практического интереса, так как такое правильное построение на практике может оказаться значительно менее точным, чем какое-нибудь простое приближенное построение. Решение этой проблемы имеет скорее принципиальное значение для истории науки. Исследования относятся к так называемым *евклидовым построениям.* Допустимыми приспособлениями здесь являются циркуль и линейка. При этом допускаются следующие приемы построения: проведение прямой через две точки; описание из данной точки окружности данного радиуса; обозначения точек пересечения двух прямых, прямой и окружности, а также двух окружностей, если они пересекаются (но не касаются!). С помощью аналитической геометрии эту задачу можно сформулировать «на языке» алгебры и установить необходимое и достаточное условие выполнимости построения. Приведем несколько классических результатов:

неразрешима задача квадратуры круга (то есть для данного круга с помощью евклидовых построений нельзя построить квадрат, площадь или периметр которого были бы равны, соответственно, площади круга или длине окружности).

Неразрешима зачача трисекции угла (то есть существуют углы, третью часть которых нельзя определить с помощью евклидовых построений; конечно, не все углы таковы).

Неразрешима задача удвоения куба (то есть из ребра данного куба при помощи евклидовых построений нельзя получить ребро куба, имеющего вдвое больший объем).

Кроме того, можно указать, какие из правильных многоугольников удается построить.

С помощью теории поля можно дать достаточно простое условие, при котором основываясь на отрезках данной длины с помощью евклидовых построений удается построить отрезок новой длины. Менее строго это условие можно сформулировать и без привлечения теории поля. *Если даны отрезки, имеющие длины $\alpha_1, \alpha_2, \ldots, \alpha_n$, то отрезок некоторой длины β можно построить (при помощи евклидовых построений) тогда и только тогда, когда β представимо в виде выражения, которое получается из величин $\alpha_1, \alpha_2, \ldots, \alpha_n$ и рациональных чисел с помощью конечного числа операций сложения, вычитания, умножения, деления и извлечения квадратного корня.*

КЛАССИФИКАЦИЯ УРАВНЕНИЙ С ОДНИМ НЕИЗВЕСТНЫМ

Уравнения с одним неизвестным обычно задаются в виде $f(x) = g(x)$, где в выражениях $f(x)$ и $g(x)$, кроме неизвестного x присутствуют также данные величины (числа, параметры или другие неизвестные). Образовывая всевозможные суммы, разности, произведения и отношения этих элементов, мы получим некоторое поле K. То, к какому типу относится рассматриваемое уравнение, определяется связью выражений $f(x)$ и $g(x)$ с кольцом многочленов $K[x]$.

Если $f(x)$ и $g(x)$ принадлежат к $K[x]$, то уравнение называется алгебраическим.

В этом случае $f(x)$ и $g(x)$ — многочлены.

Обозначим через $K(x)$ совокупность всех дробных выражений, которые могут быть образованы из многочленов кольца $K[x]$ (исключая те, которые содержат 0 в знаменателе). Очевидно, что относительно соответствующих операций $K(x)$ является полем.

Если $f(x)$ и $g(x)$ принадлежат к $K(x)$, то уравнение называется дробным алгебраическим уравнением.

Если $f(x)$ и $g(x)$ являются алгебраическими элементами над полем $K(x)$, то уравнение называется иррациональным.

Иррациональными являются, например, уравнения, в которых неизвестное присутствует под знаком радикала. Существуют, однако, и другие иррациональные уравнения. К ним относится, например, уравнение $t^2 = x + 1$, где t является корнем уравнения $t^5 - 2xt + 2 = 0$. Можно показать, что в этом уравнении неизвестное не присутствует под знаком радикала.

Если $f(x)$ и $g(x)$ являются трансцендентными элементами над полем $K(x)$, то уравнение называется трансцендентным.

(Транцендентное уравнение имеем и в том случае, если одно из выражений $f(x)$, $g(x)$ является алгебраическим элементом над полем $K(x)$).

Трансцендентными являются, например, тригонометрические, логарифмические и показательные уравнения. Это оказывается следствием того, что $\sin x$, $\cos x$ и другие, а также $\log x$ и a^x (a — параметр) являются трансцендентными элементами над полем $K(x)$.

Существуют и другие уравнения, однако, для них не используется специального названия. Такое уравнение мы получим, если одно из выражений, $f(x)$, или $g(x)$, не является ни алгебраическим, ни трансцендентным элементом над полем $K(x)$. Согласно определению алгебраического и трансцендентного элементов, это случается тогда, когда не существует такого поля L, к которому принадлежали бы $f(x)$ и $g[x]$. Примером является уравнение $|x| = x/|x|$ обозначает абсолютную величину x, то есть $|x| = x$, если $x \geq 0$, и $|x| = -x$ при $x < 0$). В этом случае $K[x]$ состоит из многочленов с рациональными коэффициентами, а $K(x)$ — из так называемых рациональных дробей. Если бы существовало расширение L поля $K(x)$, содержащее выражение $|x|$, то в поле L выполнялось бы равенство

$$(|x| + x)(|x| - x) = |x|^2 - x^2 = x^2 - x^2 = 0.$$

Поскольку в произвольном поле произведение может равняться 0 тогда и только тогда, когда хотя бы один из сомножителей равен 0, то $|x|$ должно было бы тождественно равняться x или $-x$, что не имеет места. Следовательно, указанное уравнение не относится ни к одному из перечисленных типов.

ЛОГАРИФМИЧЕСКИЕ И ПОКАЗАТЕЛЬНЫЕ УРАВНЕНИЯ

Наиболее известны три типа трансцендентных уравнений: тригонометрические, логарифмические и показательные уравнения. Уравнения первого типа рассматриваются в разделе «Геометрия». Здесь мы исследуем уравнения второго и третьего типов.

Займемся сначала показательными уравнениями, для этого нам понадобится понятие показательного выражения. По существу это выражения, в которых неизвестное может присутствовать и в показателе степени. Точнее, *показательными выражениями мы будем считать:*

1. *Рациональные выражения относительно неизвестного.* (Это целесообразно допустить по следующей причине: если $u(x)$ и $v(x)$ — рациональные выражения, то по самому виду выражения $u(x)^{v(x)}$ ясно, что оно показательное. Если же $v(x) = 1$, то получаем выражение $u(x)$. Следовательно, такие случаи нужно было бы исключить из рассмотрения. Однако сложность получающейся при этом формулировки создала бы значительно больше «неудобств», чем данное определение. Это допущение аналогично тому, что и квадрат считается прямоугольником.)

2. *Если $u(x)$ и $v(x)$ — показательные выражения, то $u(x)^{v(x)}$ также является показательным выражением.* (Благодаря сделанному выше допущению, теперь нет необходимости добавлять, что если одно из выражений $u(x)$ или $v(x)$ является показательным, а другое рациональным, то $u(x)^{v(x)}$ также является показательным и так далее.)

3. *Рациональное выражение от показательных выражений также является показательным выражением.* (Так, например, показательным является и выражение $\dfrac{2^x + x^2}{x^x - 3x}$). *Выражениями, которые получаются из рациональных выражений указанными способами, исчерпываются все показательные выражения.*

В этих трех пунктах мы, по сути дела, указали не на то, *какие* выражения являются показательными, а на то, *как их можно получить*. Такой способ определений используется главным образом в математической логике.

Уравнение $f(x) = g(x)$ называется показательным, если и $f(x)$, и $g(x)$ являются показательными выражениями.

(Если мы *непременно* хотим исключить из числа показательных уравнений алгебраические уравнения, то *теперь* к данному определению следует добавить: ... причем хотя бы одно из них не является рациональным выражением.)

Рассмотрим теперь некоторые из типов показательных уравнений, а также методы их решения. Наиболее простым является уравнение вида $a^x = b$, где a и b — положительные числа. Его решение есть $x = \log_a b$. К уравнению первой степени и уравнению предыдущего типа приводится, например, следующее уравнение:

$$2^{x+5} - 2^{x+2} - 2^x = 3^{x+2} - 3^x.$$

Используя свойства степени, получаем:

$$2^x(2^5 - 2^2 - 1) = 3^x(3^2 - 1),$$

откуда:

$$27 \cdot 2^x = 8 \cdot 3^x \qquad \text{или} \qquad 3^3 \cdot 2^x = 2^3 \cdot 3^x, \qquad \text{то есть} \left(\frac{2}{3}\right)^x = \left(\frac{2}{3}\right)^3.$$

Это уже уравнение рассмотренного выше типа, его решением является очевидно $x = 3$. При решении мы совершали такие преобразования, в результате которых «корни не могли быть потеряны» (в том числе и при делении на 3^x, так как это выражение при любых значениях x положительно). Следует, однако, проверить, является ли полученное значение корнем. Выполним подстановку: $2^8 - 2^5 - 2^3 = 256 - 32 - 8 = 216$; $3^5 - 3^x = 243 - 27 = 216$, то есть мы действительно получили решение.

Решим теперь уравнение, которое приводится к квадратному уравнению и к показательному уравнению первого типа:

$$3 \cdot 4^{x+1} + 2 \cdot 9^{x+1} + 6^x = 6^{x+2}.$$

Используя свойства степени, преобразуем уравнение таким образом, чтобы в показателе присутствовало только x:

$$2 \cdot 4 \cdot 4^x + 2 \cdot 9 \cdot 9^x + 1 \cdot 6^x = 36 \cdot 6^x.$$

Вычитая из обеих частей $1 \cdot 6^x$, а затем вынося в правой части 6^x за скобку и выполняя необходимые действия, получаем:

$$12 \cdot 4^x = 18 \cdot 9^x = 35 \cdot 6^x.$$

Если заметить, что $4 = 2 \cdot 2$, $9 = 3 \cdot 3$ и $6 = 2 \cdot 3$, то представляется целесообразным переписать уравнение в виде

$$12 \cdot (2^x)^2 + 18 \cdot (3^x)^2 = 35 \cdot 2^x \cdot 3^x.$$

Перенося все члены в левую часть, после деления на положительное выражение $(3^x)^2$, получаем:

$$12 \cdot \left[\left(\frac{2}{3}\right)^x\right]^2 - 35 \cdot \left(\frac{2}{3}\right)^x + 18 = 0.$$

Легко видеть, что здесь x присутствует лишь как показатель степени числа $\frac{2}{3}$. Следовательно, целесообразно найти сначала $y = \left(\frac{2}{3}\right)^x$. Тогда, если y уже известен, мы получим уравнение вида $a^x = b$. Для y имеем следующее уравнение:

$$12y^2 - 35y + 18 = 0.$$

Его решениями являются $\frac{9}{4}$ и $\frac{2}{3}$, которые нетрудно записать в виде степеней числа $\frac{2}{3} : \frac{9}{4} = \left(\frac{2}{3}\right)^{-2}$ и $\frac{2}{3} = \left(\frac{2}{3}\right)^1$. Таким образом для x получаются значения 1 и -2. Подставляя эти значения в исходное уравнение, получаем: если $x = 1$, то $3 \cdot 4^2 + 2 \cdot 9^2 + 6 = 48 + 162 + 6 = 216$ и $6^{1+2} = 216$; если $x = -2$, то $3 \cdot 4^{-1} + 2 \cdot 9^{-1} + 6^{-2} = \frac{3}{4} + \frac{2}{9} + \frac{1}{36} = \frac{27 + 8 + 1}{36} = 1$ и $6^{-2+2} = 1$. Таким образом оба значения являются решениями уравнения.

Прежде чем перейти и логарифмическим уравнениям, дадим определение логарифмического выражения. *Логарифмическими выражениями мы будем считать следующие:*

1. *рациональные выражения от неизвестного;*

2. *если $u(x)$ и $v(x)$ — логарифмические выражения, то $\log_{u(x)} v(x)$ также является логарифмическим выражением;*

3. *рациональное выражение от логарифмических выражений также является логарифмическим выражением;*

Выражениями, которые получаются из рациональных выражений указанными способами, исчерпываются все логарифмические выражения.

Уравнение $f(x) = g(x)$ называется логарифмическим, если и $f(x)$, и $g(x)$ являются логарифмическими выражениями.

И здесь мы рассмотрим только некоторые из типов логарифмических уравнений. Рассмотрим сначала уравнение

$$\lg(x + 6) + \lg(x - 6) = \lg(2x - 1).$$

Используя свойства логарифмов, получаем:

$$\lg(x^2-36)=\lg(2x-1).$$

Поскольку основания равны, то из равенства логарифмов следует, что $x^2-36=2x-1$. Отсюда приходим к квадратному уравнению $x^2-2x-35=0$, корнями которого являются числа -5 и 7. Число -5 не является решением исходного уравнения, так как выражение $\lg(-11)$, получаемое в левой части, не имеет смысла (в области вещественных чисел). В то же время 7 является решением, так как $\lg 13+\lg 1=\lg 13$. Основанием логарифмов здесь было число 10, однако мы этого не использовали.

Рассмотрим теперь такой пример, где существенно, по какому основанию берутся логарифмы:

$$\lg(2x^2+2)-\lg(x-6)=2.$$

Отсюда, преобразуя левую часть, получаем:

$$\lg\frac{2x^2+2}{x-6}=2.$$

Поскольку основание логарифма равно 10, а показателем (то есть величиной логарифма) является 2, то логарифмируемое выражение равно $10^2=100$, то есть $\frac{2x^2+2}{x-6}=100$. После элементарных преобразований получаем уравнение $x^2-50x+301=0$, корнями которого являются 43 и 7. Подставляя эти выражения в исходное уравнение, имеем:

$$\lg(2\cdot 1849+2)-\lg(43-6)=\lg 3700-\lg 37=2;$$

$$\lg(2\cdot 49+2)-\lg(7-6)=\lg 100=2.$$

Если аналогичным образом решить уравнение

$$\lg(x^2-1)=1+\lg(x-1),$$

то мы получим корни 9 и 1. Подстановкой можно убедиться, что 9 действительно является решением, а 1 нет, так как $\lg 0$ не имеет смысла. В то же время 1 является корнем уравнения $x^2-1=10(x-1)$, получаемого на первом шаге. Этот пример показывает, что «избавляясь» от логарифма, мы приобретаем новые корни — все те числа, которые являются корнями получаемого уравнения, но при подстановке в исходное уравнение дают логарифм нуля или отрицательного числа.

КОНЕЧНЫЕ ПОЛЯ

В заключение отметим, что в современной математике очень большое значение приобрели так называемые *конечные поля*, состоящие из конечного числа элементов. Также поля имеют весьма широкое применение.

АЛГЕБРАИЧЕСКИЕ СТРУКТУРЫ

СТРУКТУРЫ, РАССМОТРЕННЫЕ ДО СИХ ПОР

До сих пор были рассмотрены следующие алгебраические структуры: векторные пространства, группы и поля. Кроме того, мы познакомились и с кольцами многочленов. Теорию колец также можно строить аксиоматически. Под кольцом, по существу, понимается такая структура, в которой выполнимы основные арифметические действия, за исключением деления, и справедливы свойства действий. Кольцо образуют, например, линейные преобразования векторного пространства или целые числа. Это показывает, что кольца также находят широкое применение.

РЕШЕТЧАТЫЕ СТРУКТУРЫ

Можно сказать, что в рассмотренных выше структурах нас интересовали четыре основные операции, остальные же операции не принимались во внимание. Существуют, однако, алгебраические структуры совсем иного рода. В качестве примера упомянем так называемые решетчатые структуры.

Непустое множество L называется решетчатой структурой, если в нем определены две операции: пересечение, обозначаемое символом $\cap$, и объединение, обозначаемое символом $\cup$, — относительно которых выполнены следующие аксиомы:

(1)—(2) для любых двух элементов a и b существуют: их пересечение $a \cap b$ и объединение $a \cup b$

(3)—(4) операции коммутативны:

$$a \cap b = b \cap a \qquad \text{и} \qquad a \cup b = b \cup a$$

(5)—(6) операции ассоциативны:

$$(a \cap b) \cap c = a \cap (b \cap c) \qquad \text{и} \qquad (a \cup b) \cup c = a \cup (b \cup c)$$

(7)—(8) имеет место свойство «поглощения»:

$$a \cap (a \cup b) = a \qquad \text{и} \qquad a \cup (a \cap b) = a.$$

Название этих операций указывает на происхождение таких алгебраических структур. Нетрудно показать, что подмножества некоторого множества относительно пересечения и объединения в теоретико-множественном смысле образуют решетчатую структуру. Эта структура уже отличается от рассмотренных до сих пор. Бросается в глаза *двойственность* операций, то есть тот факт, что если пересечение назвать объединением и наоборот, то мы все равно получим решетчатую структуру. Операции здесь связаны не дистрибутивностью, а свойством поглощения.

Существуют, однако, решетчатые структуры особого типа, в которых имеет место и дистрибутивность. Такие решетчатые структуры называются дистрибутивными.

Для решетчатых структур можно задать «два типа» дистрибутивности:

$$(9)\ a \cap (b \cup c) = (a \cap b) \cup (a \cap c) \qquad (10)\ a \cup (b \cap c) = (a \cup b) \cap (a \cup c).$$

Здесь это не вызывает затруднений, так как можно доказать, что выполнение одной из аксиом (9) или (10) влечет за собой и выполнение другой. Особую важность имеют дистрибутивные решетчатые структуры, в которых для каждого элемента существует так называемое «дополнение». Если в решетчатой структуре существуют такие элементы 0 и e, что для любого элемента a структуры имеют место соотношения $0 \cap a = 0$ и $e \cup a = e$, то 0 называется нулевым, а e — единичным элементом структуры. Если в решетчатой структуре существуют нулевой и единичный элементы, и для элемента a найдется такой элемент a', что $a \cap a' = 0$ и $a \cup a' = e$, то a' называется *дополнением* к элементу a.

Булева алгебра. Если в дистрибутивной решетчатой структуре существуют нулевой и единичный элементы, и для любого элемента структуры найдется дополнительный элемент, то мы имеем Булеву алгебру.

Булеву алгебру образуют, например, все подмножества некоторого множества. То, что они образуют решетчатую структуру, очевидно. Нетрудно доказать и выполнение дистрибутивности. Нулевым элементом является пустое множество, а единичным — все основное множество. Для каждого подмножества существует дополнительный элемент — дополнение к множеству в теоретико-множественном смысле.

Булевы алгебры находят применение главным образом в теории множеств, в математической логике, в теории вероятностей и в функциональном анализе.

ОБЩИЕ ЗАМЕЧАНИЯ ОБ АЛГЕБРАИЧЕСКИХ СТРУКТУРАХ

Мы уже неоднократно использовали выражение «алгебраическая структура». Были приведены и примеры алгебраических структур.[1] Возникает, однако, вопрос, что вообще понимается под алгебраической структурой (то есть не под *некоторой* (конкретной) алгебраической структурой, а под алгебраической структурой *вообще*). Необходимо дать определение, охватывающее все структуры, с которыми мы познакомились и, естественно, другие важные алгебраические структуры. Если собрать все общее, что было в определениях рассмотренных структур — конечно, не всегда дословно — то получится следующее: «некоторое непустое множество называется . . ., если в нем определена . . ., операция (операции), обладающая свойствами . . .». Если теперь заполнить пропущенные места, записав на первое место название структуры, на второе — количество операций, а на третье — соответствующие свойства, то можно получить определение любой из алгебраических структур. Таким образом общее определение можно сформулировать следующим образом:

некоторые непустое множество называется алгебраической структурой, если в нем определены некоторые операции, обладающие определенными свойствами.

В связи с конкретными примерами мы уже упоминали, что алгебраическими методами нельзя, да мы и не хотим, различать алгебраические структуры, между элементами которых можно установить такое соответствие, при котором операции «ведут себя одинаково». Такие алгебраические структуры мы назвали изоморфными. Вообще: две алгебраические структуры называются изоморфными, если между ними существует взаимно однозначное соответствие, сохраняющее операцию (операции).

Теперь можно сказать, чем занимается абстрактная алгебра.

В абстрактной алгебре изучаются такие свойства отдельных алгебраических структур, которые инвариантны относительно изоморфизма.

Другими словами, можно считать, что абстрактная алгебра состоит из нескольких разделов, каждый из которых занимается изучением некоторой конкретной алгебраической структуры. При этом не принимаются во внимание ни природа элементов отдельных структур, ни названия соответствующих операций. (Например, положительные числа образуют группу относительно умножения, а вещественные числа — относительно сложения. С точки зрения алгебры между этими группами нет различия, так как они изоморфны, хотя множество положительных чисел и отличается от множества вещественных чисел. Изоморфизм здесь можно задать, поставив в соотвествие каждому положительному числу его логарифм.)

ОСНОВНЫЕ НАПРАВЛЕНИЯ СОВРЕМЕННОГО РАЗВИТИЯ АЛГЕБРЫ

Абстрактная алгебра — или классическая абстрактная алгебра — занимается изучением отдельных алгебраических структур. К этому частично относится еще одна быстро развивающаяся ветвь, называемая *теорией автоматов*.

На следующей ступени абстрагирования изучаются уже не отдельные структуры, а типы структур. Например, все векторные пространства, или все группы. Такими проблемами занимается *гомологическая алгебра*, а также теория *категорий*. В перечисленных разделах, по существу, принимаются во внимание свойства операций. В то же время в так называемой *универсальной алгебре* отвлекаются даже от этого. Еще один раздел современной алгебры, близкий к теории множеств, называется *теорией моделей*.

Многие конкретные результаты, достигнутые в отдельных разделах алгебры, могут быть использованы в других областях математики. Число таких результатов продолжает увеличиваться. Следует, однако, отметить, что пожалуй, еще большее значение, чем эти отдельные результаты, имеет то влияние, которое оказала алгебра на другие разделы математики. Немаловажную роль сыграли в этом алгебраические методы, нашедшие применение во многих разделах математики.

КРАТКИЙ ОБЗОР ИСТОРИИ АЛГЕБРЫ

Первые понятия арифметики появились еще у халдеев и финикийцев. Эти народы занимались торговлей, и развитие хозяйства привело их к необходимости выработать правила счета. Вавилонские и древнегреческие математики занимались уже и решением простейших уравнений.

Основателем алгебры считают Диофанта (409—325 г. до н. э.). Он ввел в употребление знак равенства и впервые использовал символический язык алгебры. Диофант занимался и решением систем уравнений, и неравенств первой степени.

Чуть позднее, во II и I веках до нашей эры аналогичные результаты были достигнуты и в Китае, однако, ввиду изолированности страны, они не повлияли на развитие математики.

В средние века развитие алгебры продолжили индийские математики — Ариабхата (VI век), Брамагупта (VII век) и Бхаскара (XII век). Их достижения значительно превосходит тот вклад, который внесли в развитие алгебры арабские математики. Само слово «алгебра» происходит от названия книги аль-Хорезми (IX век) «Аль-джебр аль-мукабала». В этой книге автор дает решение квадратного уравнения и занимается исследованием его корней. Он же ввел и десятичную систему счисления. Табит-ибн-Корра (834—901) перевел на арабский язык труды великих греческих философов по натурфилософии. В одной из своих работ по алгебре он поставил вопрос о решении уравнения третьей степени геометрическим путем. Алкархи (около 1000 г.) производил вычисления с иррациональными корнями. Ибн Абанна (около 1222 года преподавал в Марокко) занимался «правилами, дающими возможность вычислять неизвестное через известные числа».

Европейские алгебраисты средних веков и эпохи Возрождения также занимались вопросами, входящими ныне в курс элементарной алгебры. Формулу суммы членов геометрической прогрессии, например, связывают с именем Белдоманди (родился около 1380 года). Наиболее значительной фигурой в математике этого времени являлся Леонардо Пизанский (Фибоначчи), который родился в 1180 году. В Западной Европе алгебра и арифметика арабов стали известны благодаря его книге «Liber Abaci». Итальянцы и в дальнейшем продолжают входить в число ведущих алгебраистов. Тарталья (1500—1577) однажды за два часа решил задачи, на решение которых было дано 30 дней. Своими исследованиями он содействовал нахождению общего решения уравнения третьей степени. Кардано (1501—1576) увековечил свое имя, дав решение этого уравнения.

Крупнейшим алгебраистом эпохи Возрождения был француз Виет (1540—1603), создатель современной алгебраической символики. Свою символику он использовал при решении и преобразованиях уравнений. Занимался он и решением уравнения третьей степени. Для приближенного нахождения корней уравнений, неразрешимых алгебраическим путем, Виет впервые применил метод последовательных приближений.

Слугой, а затем учеником Кардано был Феррари (1522—1565). Он свел решение уравнения четвертой степени к решению двух квадратных уравнений.

Француз Жирар и англичанин Т. Гарриот продолжили развитие алгебраической символики (скобки, знаки $<$, $>$) и теории уравнений. Шотландский математик Непер произвел переворот в методах вычислений, открыв логарифмы; он ввел также некоторые новые важные функции. Его ученик Бригс ввел десятичные логарифмы. Используя незадолго до того открытые Непером логарифмы, Кеплер смог произвести вычисления, важные для астрономии.

Современная математика появилась в XVII веке. Первым ее представителем стал Декарт (1596—1650). Занимаясь исследованием алгебраических уравнений, он установил правило знаков, получившее его имя. Метод полной индукции впервые был применен Паскалем (1623—1662). Англичанин Уоллис (1616—1703) дал определение степени с отрицательным показателем.

Исследования Ньютона (1643—1727) относились главным образом к алгебре многочленов. Он впервые сформулировал теорему о разложении целой степени бинома. Имя Муавра (1667—1754) известно в связи с возведением в степень комплексных чисел. Швейцарский преподаватель философии Крамер (1704—1752) указал способ решения системы n уравнений первой степени с n неизвестными при помощи определителей. Известный французский энциклопедист Даламбер (1717—1783) своими исследованиями подготовил почву для доказательства основной теоремы алгебры. Эйлер (1707—1783) занимался теорией уравнений. Он обобщил теорему Ньютона на случай произвольного показателя. Он же ввел и показательную форму комплексных чисел. Безу (1730—1783) в своей книге, посвященной общей теории алгебраических уравнений, достиг значительных результатов, относящихся к роли симметрических функций от корней многочленов. Лагранж (1736—1813) при решении алгебраически разрешимых уравнений пришел к выводу о решающем значении замены переменной. Француз Вандермонд (1735—1796) занимался теорией определителей. Появление его работы о решении уравнений принято считать началом современной алгебры.

Гаусс (1777—1855) в возрасте 22 лет в докторском трактате доказал основную теорему алгебры, а затем в главном своем труде под названием «Disquisitiones Mathematicae» уделил большое внимание исследованию уравнений вида

$$x^{2n+1} = 1 \qquad (\text{где } 2n + 1 \text{ — простое число}).$$

Норвежский математик Абель (1802—1829) впервые показал, что в общем случае алгебраические уравнения выше четвертой степени неразрешимы алгебраическим путем. Французский математик Галуа (1811—1832) в ночь перед своей трагической гибелью сформулировал полученные им результаты: он установил основной фактор, определяющий все свойства алгебраических уравнений. Введенное им понятие «группы» приобрело большое значение во многих разделах математики и послужило толчком к бурному развитию алгебры.

Заслуга англичанина Гамильтона (1805—1865) — во введении кватернионов. Немецкий математик Куммер (1810—1893), занимаясь алгебраическими числовыми полями, заложил основы теории идеалов. Англичанин Сильвестер (1814—1897) занимался теорией матриц. Английский математик Кэли (1821—1895) добился значительных результатов в теории инвариантов и в теории матриц. Немецкий математик Кронекер (1823—1891) в своей работе о решении уравнения пятой степени использовал понятие группы Галуа. Дедекинд (1831—1916) продолжил развитие теории идеалов. Француз Жордан (1838—1922) с помощью разработанной им теории субституций обобщил результаты Абеля и Галуа. Большое влияние на развитие современной алгебры оказали исследования норвежского математика С. Ли (1842—1899) в теории групп.

В начале XX века была создана общая теория поля, а в двадцатых годах — теория колец и абстрактная теория групп. В тридцатых годах возникла топологическая алгебра, а в сороковых и пятидесятых годах — теория универсальных алгебр.

Из венгерских алгебраистов выделяются Ласло Редеи (1900—) и рано скончавшийся Тибор Селе (1918—1955).

ТЕОРИЯ ЧИСЕЛ

ВВЕДЕНИЕ

Теория чисел является одной из самых древних частей математики. Самое древнее математическое произведение под названием «*Элементы*» было составлено Евклидом примерно 300 лет до нашей эры. В нем составлено систематическое построение теории чисел. Вероятно в данном произведении собраны и приведены в систему результаты опыта многих поколений человечества.

К теории чисел относятся проблемы, касающиеся предмета «структурного построения» целых чисел (например, проблема разбиения целых чисел на факторы, или воспроизводимость натуральных чисел в виде суммы квадрата двух целых чисел). Основными проблемами теории чисел можно считать проблемы, возникшие при исследовании предыдущей области теории чисел, например, алгебраической теории чисел. Наконец, методы, применяющиеся в самых различных областях математики, но возникшие при решении задач по теории чисел, также могут относиться к теории чисел.

ЦЕЛЫЕ ЧИСЛА

ЭЛЕМЕНТАРНЫЕ СВОЙСТВА ЦЕЛЫХ ЧИСЕЛ

Натуральными или положительными целыми числами называются числа $1, 2, 3, \ldots$ Отрицательными числами являются натуральные числа с обратным знаком, т. е. числа $-1, -2, -3, \ldots$

Целыми числами называются положительные и отрицательные целые числа и 0.

(Неотрицательными целыми числами называются положительные целые числа и 0; аналогично подразумевается понятие совокупности неположительных целых чисел.)

Известны следующие свойства:

Существуют сложение и произведение двух целых чисел, и они являются однозначно определенными целыми числами. И для операции сложения, и для операции произведения выполняются условия *заменимости* (коммутативности) и *ассоциативности* (группируемости). Таким образом, если a, b и c представляют собой любые целые числа, то действительны соотношения:

$a+b=b+a$	и	$a\cdot b=b\cdot a$	(переместимость),
$a+(b+c)=(a+b)+c$	и	$a\cdot(b\cdot c)=(a\cdot b)\cdot c$	(ассоциативность).

Операция произведения является *дистрибутивной* (раздробляющей) в отношении сложения, т. е.:

$$(a+b)\cdot c = a\cdot c + b\cdot c.$$

Наконец, может быть осуществлено обращение операции сложения, вычитание, т. е. для любых целых чисел a и b может быть найдено такое единственное целое число x, при котором выполняется условие

$$a + x = b.$$

Напротив, обращение операции произведения, деление является неосуществимым в общем случае. Например, если умножить число 2 на любое целое число, результатом произведения никогда не может быть число 3. (В некоторых случаях и операция деления может быть осуществлена; к этому еще вернемся ниже.)

Под термином *обращения* некоторого целого числа a (или числа с отрицательным знаком) подразумевается число $-a$. Так, например, обращение числа 5 — -5, обращение числа -7 — 7, обращение 0 — 0. Существует неотрицательное значение для любого числа — а именно или само число, или его обращение, и это называется *абсолютной величиной* данного числа. На основе настоящего определения абсолютной величины некоторого числа a:

$$\text{абсолютная величина числа } a\text{: } |a| = \begin{cases} a, & \text{если } a \text{ — положительное,} \\ 0, & \text{если } a = 0, \\ -a, & \text{если } a \text{ — отрицательное.} \end{cases}$$

ДЕЛЕНИЕ С ОСТАТКОМ

При обращении операции произведения следует найти такое целое число y для заданных целых чисел a и b, в отношении которого выполяется условие $a\cdot y = b$. Если, например, $a = 5$ и $b = 20$, то таким числом является $y = 4$, так как $5\cdot 4 = 20$ (т. е. по объяснению произведения с помощью повторных сложений: $5+5+5+5 = 20$). Если изменить значение b, например, при $b = 17$, повторное сложение $5+5+5 = 15$ дает результат, меньше чем 17, но результат суммы $5+5+5+5$ больше, чем 17; следовательно, 17 больше, чем $5\cdot 3$, но отклонение меньше, чем 5; точнее: $17 = 5\cdot 3 + 2$. Аналогичным путем для любых других целых чисел a и b число b может быть записано в виде $b = a\cdot q + r$, где r является неотрицательным целым числом, меньшим, чем a. Таким образом, подобно вышеприведенному рассуждению, может быть доказано утверждение:

Для любого целого числа b и для любого целого числа a, отличающегося от 0, могут быть найдены такие целые числа q и r, для которых действительно соотношение:

$$b = a\cdot q + r.$$

(кроме того, здесь r представляет собой неотрицательное целое число, меньше, чем $|a|$).

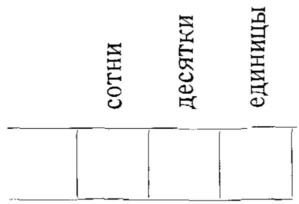
Операция, представленная выше, называется *делением с остатком*. При операции деления с остатком применяются следующие названия: *делимое* (b), *делитель* (a), *частное* (q) и *остаток* (r). Путем соответствующего доказательства можно убедиться в том, что частное и остаток деления с остатком однозначно определяются делимым и делителем. Операция деления может быть осуществлена точно, если остаток деления с остатком представляет собой 0.

СИСТЕМЫ ЧИСЕЛ

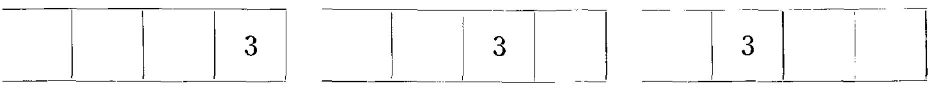
ДЕСЯТИЧНАЯ СИСТЕМА ЧИСЕЛ

Каждый человек привык манипулировать числами: осуществлять операции с ними, записывать, считая это естественным. Числа не всегда обозначались так, как в наши дни. На территории Европы до прошлого века применялись римские цифры. Маленькие числа обозначались чертами: сколько черт — такой будет величина числа: 1 = I, 2 = II, 3 = III, 4 = IIII. Для обозначения цифры 5 применялся знак V, что является символом пяти пальцев открытой руки; знак X, выражающий две кисти рук, соприкасающиеся у запястья, стоящие друг против друга, соответствовал цифре 10. Сумма чисел была записана путем применения повторных записей отдельных знаков; например, цифра 37 обозначалась в виде XXXVII. (Намного позже вычитание стало выражаться постановкой меньшего числа впереди; например, цифру 4 стали обозначать в виде IV: 4 = IV, или — 9 = IX.) Обозначение более крупных чисел исходило из их сокращенного названия. Хотя и римские цифры имеют наглядный вид, но проводить расчеты с ними трудно. Возникла потребность в развитии не только простого счета, но и счета, проводимого с большим количеством чисел и больших чисел, кроме того, быстроты исполнения операций.

Система римских цифр не могла удовлетворить таким требованиям, и весьма удобной для реализации данных целей оказалась уже оформленная к этому времени *десятичная система чисел*. Основное содержание десятичных цифр заключается в том, что в данной системе десять из каждой «единицы» переходит в новую «бóльшую единицу». Таким образом, десять единиц считается одной десяткой, десять десяток — одной сотней и т. д. Именно этим и объясняется то, что всего девять отдельных цифр требуются для выражения того, сколько берется из отдельных единиц. Но при таком обозначении нельзя сказать, обозначает ли знак 3 три единицы, 3 десятка, 3 сотни, или 3 других единиц. Но для выражения разности отдельных единиц может быть использовано положение отдельных знаков, занимаемое относительно друг друга. Расположение цифр при римских цифрах не имело специального назначения. Обозначение отдельных знаков в десятичной системе цифр осуществляется таким образом, что на последнем месте записываются единицы, на предпоследнем — десятки, а перед этим — сотни и т. д.:



Следовательно, в зависимости от того, на каком месте записывается цифра 3, разные числа будут получены:



три единицы три десятка три сотни.

Учитывая громоздкость вышеприведенной схемы, стали применять десятый знак — 0. С помощью такого знака цифры, приведенные выше в схемах, записываются так: три единицы = 3, три десятка = 30, три сотни = = 300 и т. д. В связи с этим приводятся понятия *образного значения и разряда чисел*. Например, образным значением цифры 3 является три, что является образом данной цифры и означает: сколько требуется взять из данной единицы (в данном случае — три). Разряд некоторого числа зависит от того, на каком месте оно представлено, этим определяется конкретно, из какой именно единицы следует взять три. Например, в числах 3, 30, 300 разрядом цифры 3 является единица, десяток, сотня, соответственно. Если в некоторой цифре представлено несколько цифр, это означает сумму соответствующих чисел. Например, 317 = 3 сотни + 1 десяток + 7 единиц, или: 408 = 4 сотни + 0 десяток + 8 единиц, или в более коротком виде, без указания неприсутствия десятков: 408 = 4 сотни + 8 единиц. Можно более подробно посмотреть, что действительно представляет собой некоторое число, записанное в десятичной системе чисел. Например, 54 327 обозначает число, состоящее из 5 десятков тысяч, 4 тысяч, 3 сотен, 2 десятков, 7 единиц. Поскольку каждая единица состоит из десяти единиц более маленьких единиц, единицы могут быть записаны с помощью степени числа 10: 1 сотня = 10^2 единиц, 1 тысяча = 10^3 единиц, 1 десяток тысяч = 10^4. Соответственно, вышеприведенное число может быть представлено в виде: $54\,327 = 5 \cdot 10^4 + 4 \cdot 10^3 + 3 \cdot 10^2 + 2 \cdot 10 + 7$. С помощью операции деления с остатком может быть доказано, что

каждое натуральное число может быть записано в системе десятичных цифр, т. е. в виде

$$a_n \cdot 10^n + a_{n-1} \cdot 10^{n-1} + \ldots + a_2 \cdot 10^2 + a_1 \cdot 10 + a_0,$$

где n — некоторое соответствующее неотрицательное целое число; числа

$a_n, a_{n-1}, \ldots, a_2, a_1, a_0$ — знаки цифр, т. е. неотрицательные целые числа меньше 10; $a_n \neq 0$. Вид записи любого числа десятичной системы чисел по вышеприведенному методу является однозначным, т. е. приведенные выше числа n; $a_n, a_{n-1}, \ldots, a_2, a_1, a_0$ являются однозначно определенными.

ДРУГИЕ СИСТЕМЫ ЧИСЕЛ

Если взять другое число вместо 10, то у такой системы чисел будет другая основа. Причиной того, что именно 10 единиц представляют собой новую, большую единицу, является тот простой факт, что наши руки имеют 10 пальцев. Но были разработаны и другие системы чисел. Например, в языке некоторых народов можно обнаружить *двоичную систему чисел* (в которой каждая единица состоит из *двух* более маленьких единиц). Следы *шестидесятичной системы чисел* обнаруживаются при счете времени и счете градусов углов; *двенадцатичной системы чисел* — во многих языках и в английской системе измерений. Часто системы чисел смешиваются. Например, у некоторых народностей, применяющих и большие числа, двоичная система чисел применяется только для выражения чисел меньше 10. В существовавшей английской денежной системе (1967) смешивались двенадцатичная и двадцатичная системы чисел; 12 пенни составляли 1 шиллинг и 20 шиллингов — 1 фунт. В древнем Вавилоне смешивались десятичная и шестидесятичная системы чисел; числа записывались в шестидесятичной системе чисел, но отдельные цифры были записаны в десятичной системе чисел.

Запись некоторого числа в другой системе чисел. Легко доказывается возможность записи чисел в системе чисел, основа которой отличается от оригинальной основы.

Если g представляет собой целое число больше 1, то каждое натуральное число может быть записано в «системе чисел с основой g» однозначно, т. е. в виде

$$N = b_k \cdot g^k + b_{k-1} \cdot g^{k-1} + \ldots + b_2 \cdot g^2 + b_1 \cdot g + b_0,$$

где k — неотрицательное целое число; $b_k, b_{k-1}, \ldots, b_2, b_1, b_0$ — неотрицательные целые числа, меньше g, $b_k \neq 0$.

Такая запись числа N коротко обозначается в виде

$$^{g)}b_k b_{k-1} \ldots b_2 b_1 b_0 .$$

В случае, когда $g = 10$, знак «$g)$» не записывается, т. е. вместо числа $^{10)}54\,327$ записывается просто 54 327. При этом необходимо соблюдать правило, по которому при записи числа в системе чисел с основой g «цифры» *должны быть меньше, чем g*.

Например: пусть записывается десятичная цифра 7136 ($^{10)}7136$) в шестичной системе чисел. Каждое число может быть записано в системе чисел с основой $g (g > 1)$. Сначала запишем степени 6 меньше, чем 7136:

$$6^2 = 36, \qquad 6^3 = 216, \qquad 6^4 = 1296; \qquad (6^5 = 7776 > 7136).$$

Поделим 7136 на 1296 (проводится деление с остатком!): $\begin{array}{l} 7136 : 1296 = 5, \\ \quad 656 \end{array}$

или $7136 = 5 \cdot 1296 + 656$.
Полученный таким образом остаток опять делится на самую большую степень, которая меньше, чем данный остаток, на 216: $\begin{array}{l} 656 : 216 = 3, \\ \quad 8 \end{array}$

или $656 = 3 \cdot 216 + 8$.
Если повторить данный процесс, то далее следует делить на 6 и тогда: $8 = 1 \cdot 6 + 2$. В конечном итоге:

$$7136 = 5 \cdot 1296 + 3 \cdot 216 + 1 \cdot 6 + 2 = 5 \cdot 6^4 + 3 \cdot 6^3 + 1 \cdot 6 + 2.$$

Учитывая, что вторая степень основы 6 не присутствует, т. е. присутствует 0 раз, получена следующая запись:

$$7136 = {}^{6)}53012.$$

Для предварительного счета степеней обычно требуется проводить громоздкие счеты большого объема. Но можно доказать, что и по следующему методу может быть определен вид некоторого числа в другой системе чисел: число делится на основу (проводится операция деления с остатком), частное делится опять на основу и т. д. Полученные таким образом остатки представляют собой цифры в новой системе чисел, но *в обратном порядке*. Поскольку деление проводится всегда на 6, то делитель может быть приведен впереди. Учитывая, что в таком случае уже нельзя писать знака равенства, частное отделяется от делимого чертой.

Используя приведенный выше пример, такой процесс может быть осуществлен таким образом:

$$\begin{array}{llllll} 6; & 7136/ & 1189/ & 198/ & 33/ & \underline{5} \\ & 11 & 58 & 18 & \underline{3} & \\ & 53 & 49 & \underline{0} & & \\ & 56 & \underline{1} & & & \\ & \underline{2} & & & & \end{array}$$

Если записать в обратном порядке подчеркнутые остатки: 5, 3, 0, 1, 2, то получится соответствующая последовательность цифр.

Переход в десятичную из другой системы чисел. Процесс осуществляется аналогично сделанному выше, если некоторое число, изображенное в другой системе чисел, следует переписать в десятичную систему чисел.

Например: пусть число ${}^{6)}32\,013$ записывается в десятичной системе чисел. По значению данного числа действительно соотношение:

$${}^{6)}32\,013 = 3 \cdot 6^4 + 2 \cdot 6^3 + 1 \cdot 6 + 3 = 3 \cdot 1296 + 2 \cdot 216 + 6 + 3 = 4329.$$

Данная запись также осуществляется автоматически. Все «цифры» данного числа записываются в одной строке. В следующей строке слева записывается основа, и под первой цифрой записывается сама цифра. Если данная строка заполнена до определенного места, то цифра, расположенная на следующем месте, получится таким образом, что суммируются произведения основы и числа, стоящего на предыдущем месте, и стоящее над этим местом число. Тогда число, стоящее на последнем месте, и представляет собой преобразованный вид исходного числа в десятичной системе чисел. Например, в случае, приведенном выше, это выглядит так:

	3	2	0	1	3
6	3	$20(=3\cdot 6+2)$	$120(=20\cdot 6+0)$	$721(=120\cdot 6+1)$	$4329(=721\cdot 6+3)$

ОСУЩЕСТВЛЕНИЕ РАЗЛИЧНЫХ ОПЕРАЦИЙ В СИСТЕМЕ ЧИСЕЛ

Все умеют осуществлять механически операции в десятичной системе чисел. Ответ на вопрос, почему именно таким образом проводятся расчеты, как мы привыкли, легко найти, посмотрев, как проводятся расчеты в другой системе чисел. Пусть и в этом случае основой системы чисел выбирается цифра 6, так как данная основа не является «ни слишком большой, ни слишком маленькой». Вначале рассматриваются операции, проводимые с «однозначными» числами. Этим основным операциям приходится научиться, как и таблице умножения. Полученные данные могут быть собраны в виде таблицы. В этом случае однозначными числами являются: 0, 1, 2, 3, 4, 5. Если одним членом некоторой суммы является 0, то сумма равна второму члену. Поэтому рассматриваются только другие суммы. В таблице умножения не записывается основное число, «априори» известно, что это число — 6. В первой строке и в первом столбце записываются числа; и в каждую клетку записывается сумма первых чисел, расположенных в начале данной строки и данного столбца. Получается следующая таблица:

+	1	2	3	4	5
1	2	3	4	5	10
2	3	4	5	10	11
3	4	5	10	11	12
4	5	10	11	12	13
5	10	11	12	13	14

(например, сумма 3 и 5 находится при пересечении строки с начинающей цифрой 3 и столбца с начинающей цифрой 5; $^{6)}3 + {}^{6)}5 = {}^{6)}12$.)

А сейчас посмотрим, как осуществляется суммирование многозначных чисел. Для наглядности *основное число ниже не приводится отдельно, оно остается* 6 *до конца.*

Определим сумму $315 + 543$. Пусть эти числа записываются в виде:

$$315 = 3 \cdot 100 + 1 \cdot 10 + 5 \quad \text{и} \quad 543 = 5 \cdot 100 + 4 \cdot 10 + 3.$$

(Здесь числа 10 и 100 обозначают числа в шестичной системе чисел.)

Учитывая переместимость и ассоциативность операции сложения и используя дистрибутивное свойство, получим форму:

$$315 + 543 = (3 + 5) \cdot 100 + (1 + 4) \cdot 10 + (5 + 3).$$

Подыщем значение последнего сложения из таблицы:

$$5 + 3 = 12 = 1 \cdot 10 + 2.$$

Откуда:

$$315 + 543 = (3 + 5) \cdot 100 + (1 + 4 + 1) \cdot 10 + 2.$$

Рассматривается опять последний член, содержащий операцию сложения. Эта сумма составляется на основе таблицы: $(1 + 4) + 1 = 5 + 1 = 10$; поэтому:

$$315 + 543 = (3 + 5 + 1) \cdot 100 + 0 \cdot 10 + 2.$$

Если еще раз посмотреть последний член, содержащий операцию сложения: $3 + 5 + 1 = 3 + 10 = 13 = 1 \cdot 10 + 3$, то

$$315 + 543 = 1 \cdot 1000 + 3 \cdot 100 + 0 \cdot 10 + 2 = 1203.$$

Процессом сложения доказывается, что применяется такой же метод, как и при сложении в десятичной системе чисел, поэтому операция сложения здесь также может быть осуществлена с помощью записи отдельных чисел друг под другом:

$$\begin{array}{r} 315 \\ +543 \\ \hline 1302 \end{array}$$

Процесс сложения идет таким образом: $5 + 3 = 12$ (из таблицы), записывается 2, остаток — 1. (Остаток суммируется с предыдущим столбцом) $1 + 4 = 5$; $5 + 1 = 10$, записывается 0, остаток — 1. (Переходя к столбцу, стоящему впереди) $1 + 5 = 10$; $10 + 3 = 13$, записывается 3, и — поскольку нет больше столбцов — впереди записывается 1.

Легко видеть, что и сложение большего количества членов, и операция вычитания осуществляются подобно тому, как это делается в десятичной системе чисел.

В таблице умножения представлены произведения однозначных чисел. Умножение на 0 и здесь не следует отдельно записывать, так как в случае, когда один из сомножителей 0, то и результат всего произведения будет 0. Аналогично не нужно приводить произведения, одним из сомножителей которых является 1, так как в этих случаях результат произведения всегда равняется другому сомножителю. Таблица умножения имеет следующий вид:

$\times$	2	3	4	5
2	4	10	12	14
3	10	13	20	23
4	12	20	24	32
5	14	23	32	41

(Также при пересечении соответствующей строки и соответствующего столбца можно получить результат произведения. Например: $3 \cdot 5 = 23$; действительно: $3 \cdot 5 = 5 + 5 + 5 = 5 + (1 + 4) + (2 + 3) = (5 + 1) + (4 + 2) + 3 = 10 + 10 + 3 = 23$.)

С помощью таблицы умножения произведение многозначных чисел также может быть определено.

Посмотрим, например, произведение предыдущих двух чисел. Если разложить множители:

$$315 \cdot 543 = (3 \cdot 100 + 1 \cdot 10 + 5)(5 \cdot 100 + 4 \cdot 10 + 3),$$

то на основе дистрибутивного свойства операции данное произведение разлагается таким образом:

$$315 \cdot 543 = (3 \cdot 100 + 1 \cdot 10 + 5)5 \cdot 100 + (3 \cdot 100 + 1 \cdot 10 + 5)4 \cdot 10 + (3 \cdot 100 + 1 \cdot 10 + 5)3.$$

В конечном итоге три произведения следует проводить, а затем полученные таким образом произведения суммируются. Рассмотрим сначала последний член. Используя дистрибутивное свойство и ассоциативность, переместимость операции умножения, получится следующее:

$$(3 \cdot 100 + 1 \cdot 10 + 5)3 = (3 \cdot 3) \cdot 100 + (1 \cdot 3) \cdot 10 + (5 \cdot 3).$$

Определим отдельные произведения из таблицы умножения, тогда:

$$315 \cdot 3 = 13 \cdot 100 + 3 \cdot 10 + 23 = 1300 + 30 + 23 = 1353.$$

В предпоследнем члене приходится определить результат умножения $315 \cdot 4 \cdot 10$. Определяя произведение первых двух сомножителей по вышеприведенному методу, все произведение получится как $2112 \cdot 10 = 21\,120$. Таким же образом для первого члена получится 243 100. Сложение отдельных произведений осуществляется по способу, приведенному выше:

$$\begin{array}{r} 1\,353 \\ 21\,120 \\ +243\,100 \\ \hline 310\,013 \end{array}$$

Знаки 0 в конце отдельных чисел могут быть пропущены, но оговаривается, что каждое частное произведение записывается левее на одно место. Если записать все произведение, получится следующее:

$$\begin{array}{r} 315 \cdot 543 \\ 1353 \\ \underline{2112} \\ 310013 \end{array}$$

Метод осуществляется автоматически, как и в десятичной системе чисел. Отдельно подсчитываются все «частные произведения», и затем они суммируются. Вначале осуществляется умножение на 3: $5 \cdot 3 = 23$, записывается 3, остается 2; $1 \cdot 3 = 3$, прибавляя остаток $3 + 2 = 5$, записывается и остается 0; $3 \cdot 3 = 13$ и остатка нет; поскольку это было последнее число, то записывается 13. А сейчас проводится умножение на 4: $5 \cdot 4 = 32$, записывается — 2, но на одно место левее и остается 3; $1 \cdot 4 = 4$ и прибавляя остаток: $4 + 3 = 11$, цифра 1 — стоящая на втором месте — записывается и остается 1; $3 \cdot 4 = 20$, прибавляя остаток: $20 + 1 = 21$, и так как произведение закончилось, все число записывается. Наконец, проводится умножение на 5: $5 \cdot 5 = 41$, цифра 1 — опять на одно место левее — записывается и остается 4; $1 \cdot 5 = 5$, прибавляя остаток: $5 + 4 = 13$, записывается 3, остается 1; $3 \cdot 5 = 23$, прибавляя остаток: $23 + 1 = 24$ и поскольку произведение закончено, все число записывается. После этого следует осуществить сложение частных произведений. Последняя цифра записывается без изменения. Делая шаг вперед на один знак: $5 + 2 = 11$, записывается последняя цифра 1 и остается 1. Делая шаг вперед еще раз на один знак, $3 + 1 + 1 = 5$, прибавляя остаток: $5 + 1 = 10$, записывается 0, остается 1. Шагая вперед опять на один знак: $1 + 1 + 3 = 5$, прибавляя остаток: $5 + 1 = 10$, записывается 0 и остается 1. Опять делаем шаг вперед: $2 + 4 = 10$, прибавляя остаток: $10 + 1 = 11$, записывается 1, остается 1. Наконец, прибавляя остаток к первому знаку: $2 + 1 = 3$, это записывается. Конечно, значения произведений и сложений однозначных чисел выписываются из таблиц, приведенных выше.

Деление как повторное вычитание. Операция деления может быть выполнена в виде повторного вычитания.

Пусть делится, например, число 3 440 314 на 512. (Конечно, в десятичной системе чисел.) Вместо того, чтобы брать числа 512, $2 \cdot 512, \ldots$, целесообразно рассматривать числа 512, 5120, 51 200, 512 000, 5 120 000, ... Как видно, последнее число уже больше, чем делимое, но предпоследнее — еще меньше. Поэтому число 3 440 314 делится на 512 000; итак, вычитание осуществляется не шагами в 1, а шагами в 1000. Поскольку в делителе стоит 0 на последних трех местах, то неважно, что стоит на последних трех местах делимого; проще всего будет записать туда также знаки 0. Следовательно, сейчас число 3 440 000 делится на 512 000. А вместо этого рассматривается деление 3440 : 512, так как сокращение на 1000 не меняет результата частного. Выполняя операцию деления с остатком:

$$3440 - 4 \cdot 512 = 3440 - 3252 = 144. \text{ Оттуда}:$$

$$3\,440\,314 - 4 \cdot 512\,000 = 3\,440\,000 - 3\,252\,000 + 314 =$$

$$= 144\,000 + 314 = 144\,314.$$

А сейчас уже только число 144 314 приходится делить на 512. Так как данное число меньше, чем 512 000, оно делится на 51 200; точнее, делится даже не число 144 314, а — только 144 300. Подобно предыдущему случаю, вместо настоящего деления проводится деление числа 1443 на 512. Из операции деления с остатком получится результат: $1443 - 2 \cdot 512 = 1443 - 2 \cdot 512 = 15$, то есть:

$$144\,314 - 2 \cdot 51\,200 = 144\,300 - 142\,400 + 14 = 1500 + 14 = 1514\,.$$

Сейчас число 1514 делится на 5120, точнее, 1510 делится; т. е. вместо этого число 151 делится на 512. Очевидно, результатом деления с остатком является равенство: $151 = 0 \cdot 512 + 151$. На основе данного рассуждения этот шаг описывается следующим образом:

$$1514 - 0 \cdot 5120 = 1510 - 0 + 4 = 1514\,.$$

(По существу, настоящий шаг пришлось делать только с целью обозначения деления на 5120.) Наконец, делением числа 1514 на 512 определяется равенство: $1514 = 2 \cdot 512 + 50$. Пусть записывается вся процедура. При этом не учитываются цифры, вместо которых записывался знак 0 в соответствующем шаге процедуры. Тогда получится следующая суммарная процедура деления:

$$\begin{array}{l}
3440314 : 512 = 4202 \\
\underline{-3252} \\
\,1443 \\
\underline{-1424} \\
\,151 \\
\underline{-\;0} \\
\,1514 \\
\underline{-1424} \\
50
\end{array}$$

Число, стоящее на правой стороне деления, получено следующим образом. На основе ассоциативности операции умножения вычитаемые на отдельных шагах записываются так: $512 \cdot 4000$, $512 \cdot 200$ и $512 \cdot 2$. Следовательно, частным является именно число 4202 и если всегда только первый знак записывается, то путем записи друг за другом отдельных, полученных таким образом знаков, получится именно частное.

Как было упомянуто выше, шестичная система была выбрана потому, что число 6 не слишком большое и не слишком маленькое. Если основное число слишком большое, то таблицы операций получаются слишком большие, следовательно, слишком большое количество основных операций необходимо знать наизусть. Если основное число слишком маленькое, то отдельные числа будут «слишком длинные». В таком случае 10 как основное число является довольно удачным выбором. Итак, представление операций, проводимых в шестичной системе чисел закончено; в дальнейшем основное число также будет выбрано.

ДВОИЧНАЯ СИСТЕМА ЧИСЕЛ

Как упоминалось выше, основное число некоторой системы чисел должно быть по меньшей мере 2, так как степени 1 не растут. Проведем отдельное исследование системы чисел с основным числом 2, или короче, *двоичной системы чисел*. Здесь имеются всего два знака: 0 и 1 (конечно, их можно было бы обозначать и по-другому). При составлении таблиц операций не обязательно записывать число 0 при разработке таблицы умножения и числа 0 и 1 при разработке таблицы сложения. Таким образом, *в таблицах операций двоичной системы чисел представляется лишь единственное соотношение* $1 + 1 = 10$.

Операция сложения, особенно при большом количестве членов, довольно сложно. Именно поэтому обычно только два члена суммируются одновременно. При этом операция умножения не вызывает никаких затруднений, так как, если операция умножения проводится на число 0, то ничего не нужно записывать, а если множим на 1, то следует записать множимое.

Применение системы двоичных чисел. Именно на основе простоты структуры система двоичных чисел применяется очень часто. Из возможных областей применения здесь приводятся две: 1. *электронно-вычислительные машины;* в них сигналы фиксируются в электронных лампах или в транзисторах, которые или проводят, или не проводят электрический ток в зависимости от их состояния; кроме того, сигналы могут быть зафиксированы в магнитных устройствах, намагниченность которых может иметь два направления (например, ферритовые кольца): устройство может находиться в намагниченном в или ненамагниченном состоянии (магнитный фильм). Поскольку эти устройства имеют два состояния, то с их помощью могут быть зафиксированы числа, записанные в системе двоичных чисел, и могут быть проведены расчеты с ними. (Поэтому особенно важно перечисление чисел, записанных в системе десятичных чисел, в систему двоичных чисел, или наоборот.)

Другой областью применения системы двоичных чисел является математическая логика. Пусть рассматривается утверждение A. К этому утверждению предписывается число a таким образом, что $a = 1$, если утверждение A является правильным и $a = 0$, если A ложно. На основе данного принципа логические операции, проводимые с помощью знаков *п* (правильно) и *л* (ложно), могут быть приведены к операциям, проводимым в системе двоичных чисел. Поэтому можно проводить логические операции, например, принятия различных решений, с помощью вычислительных машин. Пусть, например, к утверждению B предписывается число b. На основе утверждений, представленных выше, пусть будут правильны следующие утверждения: C = (и утверждение A, и утверждение B правильны); D = (хотя бы одно из утверждений A и B правильно). Если предписанные к утверждениям C и D числа обозначаются через c и d, соответственно, то действительны следующие соотношения: $c = a \cdot b$ и $d = a + b - a \cdot b$.

ДЕЛИМОСТЬ

Вернемся к проблеме деления с остатком, точнее, к такому частному случаю операции деления с остатком, когда остаток равен 0.

Целое число a называется делителем целого числа b, если может быть найдено такое целое число q, для которого действительно соотношение $b = a \cdot q$. Такая связь между двумя целыми числами обозначается в виде:

$$a \mid b;$$

и читается так: «число a является делителем числа b» (или: «число b делится на число a»), что обозначает, что существует некоторая связь между числом a и «как бы предварительно заданным числом b». Если число a считается предварительно заданным, и хотелось бы выразить, как число b относится к числу a, то вышеприведенный символ читается так: «число b является *кратным* числа a». Если такой связи не существует между двумя целыми числами, то это обозначается так:

$$a \nmid b;$$

что читается «число a не является делителем числа b», или «число b не является кратным числа a».

Например: $2 \mid 6$, $3 \mid -3$, $-17 \mid 51$, $-7 \mid -21$; и, $5 \nmid 7$, $-6 \nmid 2$, $8 \nmid -9$, $-5 \nmid -3$ и т. д.

В проблеме делимости особенно важным является условие, по которому число q *также* должно быть целым числом. Если, например, было бы допущено применение любого рационального числа (см. гл. «Алгебра»), то каждое число, отличающееся от 0, было бы делителем каждого числа. В общем случае от числа q требуется «иметь такой же характер», как и данные числа. Если, например, рассматриваются только четные числа (в отношении этих чисел действительны те же основные характеристики, что и в отношении целых чисел), то и для числа q допускаются только четные числа. Так, в области четных чисел 2 не является делителем числа 6 (так как число 3 — нечетное), но оно же является делителем числа 8, так как 4 — четное число.

Основные свойства делимости:

1. В отношении делимости каждое число характеризуется так же, как и его обратное; иными словами, условия $a \mid b$, $(-a) \mid b$, $a \mid (-b)$ и $(-a) \mid (-b)$ являются эквивалентными (равноценными). (Из выполнения любого приведенного условия вытекает и выполнение других трех условий.)

2. В отношении любого целого числа a — $a \mid 0$; если для некоторого целого числа b — $0 \mid b$, то $b = 0$.

Здесь нельзя путать понятия операции деления и делимости. А ведь число 0 не делится на 0, так как результат деления неоднозначен. С другой стороны, $0 \mid 0$, так как при делимости не учитывается требование однозначности числа «q» (например, $0 = 0 \cdot 1$, или $0 = 0 \cdot (-193)$).

3. В отношении любого целого числа b — $1 \mid b$, и если для некоторого целого числа a — $a \mid 1$, то $a = 1$ (или $a = -1$).

4. Для любого целого числа $a \mid a$ (длеимость является рефлексивной).

5. Если в отношении целых чисел a, b, c выполняются условия $a \mid b$ и $b \mid c$, то действительно и соотношение $a \mid c$ (делимость является транзитивной).

6. Если в отношении целых чисел a и b — $a \mid b$ и $b \mid a$, то $a = b$ или $a = -b$ (делимость является антисимметричной).

КЛАССИФИКАЦИЯ ЦЕЛЫХ ЧИСЕЛ НА ОСНОВЕ ДЕЛИМОСТИ

Учитывая, что с точки зрения делимости все числа характеризуются теми же свойствами, что и их обратные числа, будет достаточно рассматривать только неотрицательные числа и положительные делители.

Целые числа классифицируются в зависимости от того, сколько делителей они имеют.

Числа, имеющие единственный положительный делитель, называются единицами.

Таким числом является 1. Среди всех целых чисел число -1 также является единицей. Нетрудно убедиться в том, что некоторое число представляет собой единицу тогда и только тогда, если оно является делителем всех целых чисел.

Числа, имеющие два положительных делителя, называются неделимыми числами (простыми числами) или неразложимыми числами.

Такие числа: 2, 3, 5, ...; среди всех целых чисел неделимыми числами являются также -2, -3, -5, ... По существу, свойство неразложимости является важным свойством. Можно показать, что некоторое целое число тогда и только тогда является неделимым, если при его разложении на произведение двух сомножителей, один из сомножителей — единица.

Числа, имеющие ограниченное количество положительных делителей, количество которых больше, чем два, называются сложными числами.

Только число 0 имеет бесконечное количество делителей.

НАИБОЛЬШИЙ ОБЩИЙ ДЕЛИТЕЛЬ И НАИМЕНЬШЕЕ ОБЩЕЕ КРАТНОЕ

При операциях, проводимых с дробными числами, часто применяется такое свойство, по которому значение дроби не меняется, если и числитель, и знаменатель дроби перемножается или делится на одно и то же число. Последний случай появляется при сокращении дробей, когда целесообразно деление проводить на число, имеющее «как можно большее» значение; а предыдущий случай (умножение) применяется в ходе процесса приведения к одному знаменателю: здесь целесообразно найти «как можно меньший» общий знаменатель. Отсюда появляются понятия наибольшего общего делителя и наименьшего общего кратного. Можно показать, что среди общих делителей двух или больше целых чисел всегда существует самый большой делитель, а среди общих кратных — самый маленький.

Целое число d называется наибольшим общим делителем целых чисел a и b, если d является общим делителем чисел a и b, и d является кратным каждого общего делителя чисел a и b.

Так как с точки зрения делимости каждое обратное число имеет такие же свойства, как и оригинальное число, то на основе определения, приведенного выше, понятие наибольшего общего делителя стало бы неоднозначным. Чтобы данное определение было однозначным, ставится условие *положительности наибольшего общего делителя по определению.*

Наибольший общий делитель (в дальнейшем в сокращенном виде: НБОД) чисел a и b обозначается так:

$$\boxed{d = (a, b).}$$

Из приведенного выше определения еще не видна однозначность понятия НБОД, не ясно, что любые два числа имеют НБОД. В однозначности НБОД нетрудно убедиться: если и число d, и число d', являются НБОД чисел a и b, то на основе определения $d|d'$, и $d'|d$. А поскольку оба числа d и d', являются положительными, то из асимметричности делимости вытекает правильность равенства $d = d'$.

Трудно доказать, что действительно существует НБОД любых двух чисел. Как мы уже видели, при исследовании делимости отрицательные числа могут быть исключены. В случае, когда $a = b = 0$, из определения вытекает соотношение $(a, b) = 0$. Поэтому вместо положительности НБОД кажется целесообраазным ставить условие *неотрицательности НБОД*. Если одно из двух чисел, например $b = 0$ и $a \neq 0$, то из определения вытекает правильность соотношения $(a, b) = (a, 0) = a$. Поэтому можно предполагать, что и число a, и число b являются положительными.

Существование НБОД можно доказать следующим образом. Рассматривается множество чисел вида $a \cdot x + b \cdot y$, где a и b представляют собой заданные числа. При этом x и y принимают значения всех целых чисел, независимо друг от друга. Можно показать, что наименьшим положительным элементом настоящего множества является (a, b).

ОПРЕДЕЛЕНИЕ НБОД С ПОМОЩЬЮ ЕВКЛИДОВОГО АЛГОРИТМА

Доказательство существования НБОД с помощью подхода, приведенного выше, дает возможность одновременно и для конкретного определения НБОД двух заданных чисел, но данный метод неприемлем для практики. Поэтому ниже представлен такой метод определения НБОД, с помощью которого НБОД любых двух чисел может быть определен в результате ограниченного количества шагов. Методы, с помощью которых желаемый результат определяется, независимо от данных, при ограниченном количестве шагов расчета, называются алгоритмами.

Алгоритмом является, например, процедура деления, преподаваемая в школах.

Конечно, не каждая процедура является алгоритмом. Пусть, например, рассматривается натуральное число n и записывается вид десятичной дроби $\frac{1}{n}$; всегда можно определить, на каком месте после десятичного знака появится цифра 1 в первый раз. Но определение данного места всегда зависит от числа n.

Евклидов алгоритм. Евклидов алгоритм основывается на выявлении следующего легко доказуемого правила: если $a = b \cdot q + r$, то $(a, b) = (b, r)$. Остаток деления с остатком всегда меньше, чем делитель (при операции положительными целыми числами), поэтому на основе приведенного выше рассуждения задача нахождения НБОД двух чисел может быть приведена к задаче нахождения НБОД двух меньших чисел.

Определим, например, НБОД чисел 360 и 225:

$$\begin{aligned} 360 &= 225 \cdot 1 + 135; \text{ и } (360,\ 225) = (225,\ 135) \\ 225 &= 135 \cdot 1 + 90; \text{ и } (225,\ 135) = (135,\ 90) \\ 135 &= 90 \cdot 1 + 45; \text{ и } (135,\ 90) = (90,\ 45) \\ 90 &= 45 \cdot 2 + 0; \text{ и } (90,\ 45) = 45. \end{aligned}$$

То есть, $(360,\ 225) = 45$.

Так как в подобных случаях в остатке оказываются неотрицательные целые числа, и они все уменьшаются, то процедура заканчивается за ограниченное число шагов. НБОД определяется последним остатком, отличающимся от 0. Путем применения некоторых модификаций евклидов алгоритм может быть «ускорен».

Когда НБОД двух чисел является 1, то эти два числа называются относительными неделимыми числами в отношении друг к другу.

Подобно определению НБОД может быть введено понятие *наименьшего общего кратного* двух чисел (НМОК).

Неотрицательное число c называется НМОК целых чисел a и b, если число c является общим кратным чисел a и b и оно является делителем всех общих кратных этих же двух чисел. НМОК целых чисел a и b обозначается символом $[a,\ b]$.

НМОК определяется с помощью следующего соотношения:

$$\boxed{[a,\ b] \cdot (a, b) = a \cdot b.}$$

Например: $[360,\ 225] = 360 \cdot 225 = 3240.$

О НЕДЕЛИМЫХ ЧИСЛАХ

ОСНОВНАЯ ТЕОРЕМА ТЕОРИИ ЧИСЕЛ

Вычисление как НБОД, так и НМОК проводится проще, если числа разлагаются на сомножители. Предположим, нужно определить НБОД чисел a и b, и данные числа имеют вид $a = c \cdot a_1$; $b = c \cdot b_1$. В таком случае непосредственно на основе определения НБОД может быть показано, что действительно соотношение $(a,\ b) = c \cdot (a_1,\ b_1)$. Так, например, $(3060,\ 7020) = (10 \cdot 306,\ 10 \cdot 702) = 10 \cdot (306,\ 702)$. Вычисление НБОД также упрощается, если учесть утверждение: если при вычислении НБОД двух чисел одно из чисел имеет сомножитель, являющийся относительным неделимым числом по отношению к другому, то данный сомножитель может быть пропущен, т. е. если $b = u \cdot v$ и $(a,\ u) = 1$, то $(a,\ b) = (a,\ v)$.

Например, числа 317 и 10 являются относительными неделимыми числами, поэтому: $(317, 510) = (317, 10 \cdot 51) = (317, 51)$.

На основе такого рассуждения было бы целесообразно разложить числа на произведения нескольких сомножителей. При определении НБОД общие сомножители могут быть вынесены, а относительные неделимые сомножители по отношению к другому сомножителю могут быть пропущены. Поскольку, как мы уже видели, при обсуждении вопросов, связанных с делимостью, знаком чисел можно не считаться, достаточно рассматривать только положительные числа. (Также не рассматривается и число 0, поскольку оно не представляет интерес в данном случае). Мы уже видели, что неделимые числа являются неразложимыми; любое сложное число разлагается на два сомножителя, каждый из которых меньше, чем само число (поэтому такие числа называются сложными). Если какой-то из этих сомножителей является сложным числом, то он также разлагается на произведение двух сомножителей, каждый из которых меньше, чем он (и оба положительные) и т. д. Каждый из полученных таким образом сомножителей является делителем исходного заданного числа, и так как некоторое сложное число имеет ограниченное число делителей, то данная процедура заканчивается за ограниченное число шагов. А это означает, что ни один из полученных таким образом сомножителей не представляет собой сложное число и поскольку каждый из сомножителей является положительным числом, отличающимся от 1, то каждый сомножитель представляет собой неделимое число. В конечном итоге оказывается, что каждое сложное число разлагается на произведение неделимых чисел. Если при разложении для однородности рассуждения учитываются и виды с единственным сомножителем (например, такие как $2 = 2$ и т. д.), тогда вышеприведенный результат имеет силу и в отношении неделимых чисел. Следовательно, получен результат, по которому:

Каждое число, большее чем 1, может быть представлено в виде произведения неделимых чисел.

Покажем, что данное разложение является однозначным. Под однозначностью здесь подразумевается, что два разложения могут различаться только в отношении последовательности сомножителей, т. е. в обоих разложениях присутствуют одни и те же неделимые числа, и отдельные неделимые числа присутствуют столько же раз в обоих разложениях.

Здесь воспользуемся важным свойством неделимых чисел:

Если некоторое неделимое число является делителем некоторого произведения, то оно является и делителем хотя бы одного из сомножителей произведения.

Это доказывается для двух сомножителей, в случае существования нескольких сомножителей доказательство осуществляется в нескольких шагах. Предположим, что неделимое число p является делителем произведения $a \cdot b$. По определению $(p, a)|p$, и так как p представляет собой неделимое число, то или $(p, a) = p$, или $(p, a) = 1$. В первом случае, на основе определения НБОД — $p|a$, следовательно, доказательство закончено. Во втором случае, в соответствии с упомянутым выше свойством НБОД: $(p \cdot b, a \cdot b) = b \cdot (p, a) = b \cdot 1 = b$. Поскольку по предположению — $p|a \cdot b$ и $p|p \cdot b$, то $p|(p \cdot b, a \cdot b)$, т. е. $p|b$. Утверждение доказано.

Предположим, что некоторое число разложено по двум различным способам на произведение неделимых чисел:

$$p_1 \cdot p_2 \cdot \ldots \cdot p_r = q_1 \cdot \cdot q_2 \cdot \ldots \cdot q_s.$$

Неделимое число p_1 — делитель левой стороны, а потому является делителем и правой стороны. Но если некоторое неделимое число является делителем произведения, то оно должно быть делителем и некоторого сомножителя произведения, например $p_1|q_1$. Учитывая, что и число q_1 является неделимым числом, а потому имеет всего только два (положительных) делителя и поскольку число p_1 отличается от 1, то может быть только $p_1 = q_1$. Теперь можно сократить на этот сомножитель, и процедура продолжается исследованием следующего неделимого числа. Оно также должно равняться некоторому сомножителю неделимого числа правой стороны, например $p_2 = q_2$. После r-го шага получится соотношение:

$$p_r = q_r$$

(путем соответствующей перегруппировки сомножителей правой стороны), и на левой стороне остается 1, поэтому и на правой стороне должно остаться число 1. Таким образом доказано вышеприведенное утверждение, являющееся *основной теоремой теории чисел.*

Каноническая форма чисел. При записи одинаковых неделимых сомножителей в виде степени с общим основным числом получится:

Каждое целое число, большее 1, может быть записано однозначно в форме

$$\boxed{n = p_1^{\alpha_1} \cdot p_2^{\alpha_2} \cdot \ldots \cdot p_r^{\alpha_r},}$$

где $p_1, p_2, \ldots, p_r$ — различные неделимые числа и $\alpha_1, \alpha_2, \ldots \alpha_r$ — натуральные числа.

Данная форма представления чисел называется канонической формой представления чисел.

В каноническом представлении некоторого числа экспонентом отдельных неделимых чисел обозначается, сколько раз данное неделимое число встречается как сомножитель в разложении числа на произведение неделимых сомножителей. Часто целесообразно учитывать и неделимые числа, не присутствующие в разложении числа на произведение неделимых сомножителей. Экспонентом этих чисел является 0. Конечно, таким образом уже нет однозначности разложения; например, $18 = 2 \cdot 3^2 = 2 \cdot 3^2 \cdot 17^0$. Но, с другой стороны, большим преимуществом такого вида является возможность непосредственного разложения из разложения сомножителей; если

$$a = p_1^{\alpha_1} \cdot p_2^{\alpha_2} \cdot \ldots \cdot p_r^{\alpha_r} \text{ и } b = p_1^{\beta_1} \cdot p_2^{\beta_2} \cdot \ldots \cdot p_r^{\beta_r},$$

где экспоненты — неотрицательные целые числа, тогда:

$$a \cdot b = p_1^{\alpha_1+\beta_1} \cdot p_2^{\alpha_2+\beta_2} \cdot , \ldots \cdot p_r^{\alpha_r+} .$$

Пример: $a = 168$ и $b = 297$; путем представления чисел в виде $a = 2^3 \cdot 3 \cdot \cdot 7$, $b = 3^3 \cdot 11$ и при записи экспонентов 0 также: $a = 2^3 \cdot 3 \cdot 7 \cdot 11^0 \cdot$, $b = 2^0 \cdot 3^3 \cdot 7^0 \cdot 11$, и, таким образом, $a \cdot b = 2^3 \cdot 3^4 \cdot 7 \cdot 11$.

Также полезно и такое представление при вычислении НБОД и НМОК в отношении чисел a и b : $d = (a, b)$ и $c = [a, b]$.

Прежде чем переходить к представлению такой формы, приведем два обозначения. Пусть α и β представляют собой два (не обязательно целых) числа. Пусть $\min(\alpha, \beta)$ обозначает минимум этих двух чисел, т. е. меньшее из двух, а если они одинаковые, то их общее значение. Аналогично, пусть $\max(\alpha, \beta)$ обозначает максимум этих двух чисел, т. е. самое большое из них, а если они имеют одинаковое значение, то — их общее значение. Например, $\min(5,7) = 5$ и $\max(5,7) = 7$; или $\min(-2, -5) = -5$ и $\max(-2, -5) = -2$; или $\min(8,8) = 8$ и $\max(8,8) = 8$. Итак, с помощью таких обозначений НБОД и НМОК могут быть выражены в виде:

$$d = p_1^{\delta_1} \cdot p_2^{\delta_2} \cdot \ldots \cdot p_r^{\delta_r} \quad \text{и} \quad c = p_1^{\gamma_1} p_2^{\gamma_2} \cdot \ldots \cdot p_r^{\gamma_r} ,$$

где $\delta_i = \min(\alpha_i, \beta_i)$ и $\gamma_i = \max(\alpha_i, \beta_i)$ для каждого натурального числа i, имеющего значение между границами 1 и r.

Пусть определяется, например, НБОД и НМОК чисел 168 и 20 250! $168 = 2^3 \cdot 3 \cdot 7$ и $20\,250 = 2 \cdot 3^4 \cdot 5^3$, откуда: $(168, 20\,250) = 2 \cdot 3 = 6$ и $[168, 20\,250] = 2^3 \cdot 3^4 \cdot 5^3 \cdot 7 = 567\,000$.

С применением канонической формы представления чисел и дробные числа могут быть записаны в виде канонической формы, но в этом случае экспоненты могут быть и отрицательные целые числа. Определенная однозначность действительна и для канонического представления дробных чисел.

КОЛИЧЕСТВО НЕДЕЛИМЫХ ЧИСЕЛ

Мы видели, что каждое целоее число, больше 1, может быть разложено на произведение неделимых чисел. Следовательно, зная неделимые числа, все числа могут быть записаны. Но задать все неделимые числа так же невозможно, как и задать все целые числа, так как *существует бесконечное количество неделимых чисел.*

Доказательство. О нем уже говорилось в цитированной выше книге Евклида и оно представляет собой самое простое доказательство того, что количество неделимых чисел бесконечно. Для того, чтобы показать бесконечность неделимых чисел, необходимо доказать, что какими бы ни были заданные неделимые числа, всегда могут быть найдены неделимые числа, отличающиеся от заданных.

Предположим, что $p_1, p_2, \ldots, p_n$ — заданные неделимые числа. Покажем, что существует неделимое число, *отличающееся* от заданных. Пусть к произведению заданных неделимых чисел прибавится единица. Полученное таким образом число N можно представить в виде:

$$N = p_1 \cdot p_2 \cdot \ldots \cdot p_n + 1.$$

Итак, данное число N будет наверняка больше, чем 1, таким образом, существует его неделимый делитель (т. е. делитель, являющийся неделимым числом). *Пусть один из этих неделимых делителей обозначается через* p (если случайно само число N является неделимым числом, то $p = N$). На этом неделимом числе p покажем, что оно отличается от всех заданных неделимых чисел. Пусть число N имеет вид $N = p \cdot N_1$. Если число p равнялось бы некоторому числу p_1 из заданных неделимых чисел, то было бы получено соотношение:

$$1 = N - p_1 \cdot p_2 \cdot \ldots \cdot p_n = p \cdot (N_1 - p_2 \cdot \ldots \cdot p_n),$$

что исключается, так как число 1 не делится ни на одно неделимое число.

1. Метод доказательства, примененный здесь, называется *косвенным доказательством.* При использовании косвенного доказательства правильность некоторого утверждения показывается таким образом, что доказывается следующее: отрицание утверждения приводит к противоречию.

2. В ходе доказательства мы не могли задать конкретно новое неделимое висло. Удалось только показать, что *существует* неделимое число, отличающееся от заданных. Если в ходе доказательства показывается существование какого-то предмета, оно называется *доказательством существования.*

3. Если доказывается не только существование некоторого предмета, но данный предмет задается, конкретно конструируется, то это *конструктивное доказательство.* Конструктивное доказательство применялось, например, когда было показано, что любые два натуральных числа имеют НБОД.

Распределение неделимых чисел. Как мы видели, существует бесконечное количество неделимых чисел. Посмотрим, каким образом неделимые числа расположены среди целых чисел. Если подсчитать количество неделимых чисел до 10 000, то окажется, что

до 10 имеются 4 неделимых числа (40% от всех чисел до настоящей границы);

до 100 имеются 25 неделимых чисел (25% от всех чисел до настоящей границы);

до 1000 имеются 168 неделимых чисел (16,8% от всех чисел до настоящей границы),

до 10 000 имеются 1239 неделимых чисел (12,4% от всех чисел до настоящей границы).

Таким образом, если мы дойдем до большей границы, то найдется относительно меньшее число неделимых чисел. Если это действительно так, то неделимые числа должны падать «на все растущую дистанцию друг от друга». В некотором смысле это правильно; доказательство Евклида было аналогичным. В частности, если задается натуральное число n любой величины, всегда могут быть найдены n чисел, расположенных непосредственно друг за другом, ни одно из которых не является неделимым числом. Принцип доказательства подобен предыдущему. Сейчас рассматривается случай, когда $n = 10$; в общем случае доказательство проводится аналогичным образом.

Если бы нужно было показать, что существует 10 целых чисел, расположенных непосредственно друг за другом, ни одно из которых не является неделимым числом, то необходимо взять произведение всех неделимых чисел до границы 11 $(10+1)$; оно составляет $2 \cdot 3 \cdot 5 \cdot 7 \cdot 11 = 2310$. Затем, к полученному таким образом числу прибавляются числа 2, 3, 4, ... 11 соответственно. Таким образом, получены числа: 2312, 2313, 2314, 2315, 2316, 2317, 2318, 2319, 2320 и 2321. Покажем, что ни одно из этих чисел не может быть неделимым. Пусть берется один неделимый делитель (например, наименьший) каждого из чисел от 2 до 11. Данный неделимый делитель является делителем и сформированного произведения и поэтому может быть вынесен из соответствующей суммы. В рассматриваемом случае:

$$\begin{aligned}
2310 + 2 &= 2 \cdot (3 \cdot 5 \cdot 7 \cdot 11 + 1) \\
2310 + 3 &= 3 \cdot (2 \cdot 5 \cdot 7 \cdot 11 + 1) \\
2310 + 4 &= 2 \cdot (3 \cdot 5 \cdot 7 \cdot 11 + 2) \\
2310 + 5 &= 5 \cdot (2 \cdot 3 \cdot 7 \cdot 11 + 1) \\
2310 + 6 &= 2 \cdot (3 \cdot 5 \cdot 7 \cdot 11 + 3) \\
2310 + 7 &= 7 \cdot (2 \cdot 3 \cdot 5 \cdot 11 + 1) \\
2310 + 8 &= 2 \cdot (3 \cdot 5 \cdot 7 \cdot 11 + 4) \\
2310 + 9 &= 3 \cdot (2 \cdot 5 \cdot 7 \cdot 11 + 3) \\
2310 + 10 &= 2 \cdot (3 \cdot 5 \cdot 7 \cdot 11 + 5) \\
2310 + 11 &= 11 \cdot (2 \cdot 3 \cdot 5 \cdot 7 + 1)
\end{aligned}$$

Следовательно, стоящее в левой стороне число во всех случаях делится на первый сомножитель правой стороны. Но этот сомножитель представляет собой неделимое число и поэтому необходимо показать, что число, стоящее на левой стороне, не равняется этому неделимому числу, т. е. что второй сомножитель правой стороны больше 1. Но, с другой стороны, это очевидно, так как данный сомножитель представляет собой двухчленную сумму, оба члена которой составляют по меньшей мере 1 и поэтому их сумма составляет по меньшей мере 2. Конечно, таким образом получится не «самый первый разрыв с длиной в 10». Например, из 13 чисел, расположенных между числами 113 и 127, ни одно не представляет собой неделимое число.

СИТО ЭРАТОСТЕНА

На основе рассуждения, приведенного выше, неделимые числа расположены довольно «произвольно» среди целых чисел. Для нахождения всех неделимых чисел до некоторой достаточно большой границы существует так называемое *сито Эратостена*, названное так оттого, что при использовании данного метода все сложные числа «просеиваются», а оставшиеся представляют собой неделимые числа. Попробуем определить, например, все неделимые числа до 10.

Запишем все натуральные числа от 1 до 100. Первое число — 2 — большее 1, является неделимым числом. Но среди кратных 2 уже не найдется ни одного неделимого числа. Пусть остается 2 (обозначим это, например, кругом) и вычеркивается (просеивается) каждое второе число, начиная с этого числа. Первым невычеркнутым числом является 3 — опять неделимое число. Оно также остается (обводится кругом) и, начиная с этого числа, опять просеивается, вычеркивается каждое третье число, поскольку каждое из них делится на 3 и поэтому не может представлять собой неделимое число. Следующее невычеркнутое число 5 является также неделимым; оно и должно быть неделимым числом, т. к. в противном случае мы бы его вычеркнули уже тогда, когда были вычеркнуты кратные его наименьшего неделимого делителя. Настоящим рассуждением одновременно дается и объяснение способа просеивания. Ведь из приведенного рассуждения видно, что после каждого «шага просеивания» первое невычеркнутое число должно представлять собой неделимое число, и в следующем шаге необходимо просеивать кратные этих неделимых чисел. Единственным вопросом остается один: до какой границы следует проводить просеивание. Рассмотрим это дальше на вышеприведенном примере. Если кратные 5 уже вычеркнуты, то следующим числом является 7. После вычеркивания кратных 7, следующим числом является 11. Если мы хотим вычеркивать кратные 11, то оказывается, что они уже все вычеркнуты, т. е. вычеркнуты все сложные числа; «остались на сите» только неделимые. Если числа были бы записаны в таблицу, то в конечном итоге получилась бы следующая картина:

1	2	3	~~4~~	5	~~6~~	7	~~8~~	~~9~~	~~10~~
11	~~12~~	13	~~14~~	~~15~~	~~16~~	17	~~18~~	19	~~20~~
~~21~~	~~22~~	23	~~24~~	~~25~~	~~26~~	~~27~~	~~28~~	29	~~30~~
31	~~32~~	~~33~~	~~34~~	~~35~~	~~36~~	37	~~38~~	~~39~~	~~40~~
41	~~42~~	43	~~44~~	~~45~~	~~46~~	47	~~48~~	~~49~~	~~50~~
~~51~~	~~52~~	53	~~54~~	~~55~~	~~56~~	~~57~~	~~58~~	59	~~60~~
61	~~62~~	~~63~~	~~64~~	~~65~~	~~66~~	67	~~68~~	~~69~~	~~70~~
71	~~72~~	73	~~74~~	~~75~~	~~76~~	~~77~~	~~78~~	79	~~80~~
~~81~~	~~82~~	~~83~~	~~84~~	~~85~~	~~86~~	~~87~~	~~88~~	89	~~90~~
~~91~~	~~92~~	~~93~~	~~94~~	~~95~~	~~96~~	97	~~98~~	~~99~~	~~100~~

Очевидно, что здесь оказалось достаточным просеивание кратных неделимых чисел, расположенных в первом ряду. Это естественно, так как если число, меньшее 100, разлагается на произведение двух сомножителей, то хотя бы один из сомножителей должен быть меньше, чем 10. Таким образом, если число неделимое, то его наименьший неделимый делитель будет наверняка меньше 10; следовательно, уже при вычеркивании кратных некоторого неделимого числа, которое меньше, чем 10, надо было бы просеить данное число.

В общем случае, если определить неделимые числа до N, то достаточно просеивать кратные неделимых чисел до границы $\sqrt{N}$.

В отношении самого наименьшего неделимого делителя имеет место неравенство $p \leqq \sqrt{n}$, на основе которого могут быть приведены еще и другие выводы. Если следует принять решение, является ли некоторое натуральное число n неделимым, достаточно рассмотреть вопрос, имеет ли оно делитель среди неделимых чисел до границы $\sqrt{n}$. Если такой делитель не существует, то n обязательно будет неделимым числом (если есть такой делитель, то число, видимо, сложное).

ОЦЕНКИ ДЛЯ НЕДЕЛИМЫХ ЧИСЕЛ

Как мы уже видели, неделимые числа расположены «все более редко», хотя и существует бесконечное количество неделимых чисел; бывает, что в некоторых местах неделимые числа «сгущаются». Мы видели также, что все меньше процентов от всех чисел до границ 10, 100, 1000, 10 000 являются неделимыми числами. Возникает вопрос: какой процент всех натуральных чисел составляют неделимые. В некоторых случаях легко дать ответ на вопрос, какая часть чисел имеет определенное свойство (например, одна половина натуральных чисел является четной, другая — нечетной). Но если бы рассматривались не все натуральные числа, то ответ уже не был бы настолько однозначным. До некоторого четного числа действительно половина чисел является четной, но если исследование проводить до некоторого нечетного числа, то положение изменится. Если n является нечетным числом, то до числа n имеются $\frac{n-1}{2}$ четных чисел, что представляет собой часть исследуемых чисел, получаемую по соотношению $\frac{n-1}{2} : n = \frac{n-1}{2} = \left(\frac{1}{2} - \frac{2}{n}\right)$. Но если число n достаточно велико, то данная величина будет «примерно» 1/2. Точнее говоря, при n, стремящемся к бесконечности, граничным значением данного выражения является 1/2.

Таким же образом можно поступить при оценке количества неделимых чисел. Количество неделимых чисел до границы n (включая само число n, если оно является неделимым числом) обозначается символом $\pi(n)$. В таком случае часть чисел, являющаяся совокупностью неделимых чисел, может быть определена путем образования граничного значения $\lim\limits_{n \to \infty} \left(\frac{\pi(n)}{n}\right)$.

(относительно граничных значений см. главу «Анализ».) Может быть доказано, что настоящее граничное значение равняется 0, т. е. в прогрессии натуральных чисел неделимые числа действительно расположены все более редко. Тот факт, что вышеприведенное граничное значение есть 0 еще не достаточен, чтобы знать, как расположены неделимые числа. Например, пусть $N(n)$ обозначает количество квадратных чисел не больше n, и $H(n)$ — степени 2, которые не больше n. Очевидно, что $\sqrt{n} - 1 < N(n) \leq \sqrt{n}$ и $\overset{2}{\smile}\!\!\log(n) - 1 < H(n) \leq \overset{2}{\smile}\!\!\log(n)$. Из этого вытекает, что как $\lim\limits_{n\to\infty} \frac{N(n)}{n} = 0$, так и $\lim\limits_{n\to\infty} \frac{H(n)}{n} = 0$. Следовательно, граничное значение, соответствующее исследуемому граничному значению, равно 0 в обоих случаях, но количество степеней 2 «намного меньше», чем количество квадратных чисел. Выше было отмечено, что количество степеней 2 до границы n составляет не больше $\overset{2}{\smile}\!\!\log n$, которое при достаточно большом значении n существенно меньше, чем значение $\sqrt{n} - 1$, которое в свою очередь меньше, чем количество квадратных чисел до границы n. Именно поэтому было бы целесообразно сформулировать подобное соотношение и в связи с формулой $\pi(n)$. Нетрудно показать, что существуют такие постоянные значения c_1 и c_2 $(0 < c_1 < 1 < c_2)$ для которых действительно неравенство:

$$c_1 \cdot \frac{n}{\log n} < \pi(n) < c_2 \cdot \frac{n}{\log n}.$$

Настоящее неравенство имеет смысл только при $n > 1$, но это не затрудняет положение, так как каждое неделимое число больше 1. (В области теории неделимых чисел обозначение «log» указывает на применение натуральных логарифмов.)

Настоящее неравенство считается полезными и потому, что с его помощью функция $\pi(n)$ попадает в такую «полосу», границы которой определяются двумя кратными функции, описываемой аналитически довольно хорошо. В то же время еще неизвестно, как функция $\pi(n)$ характеризуется внутри данной полосы. Не исключена возможность, что значения данной функции колеблются между значениями двух граничных функций, но возможно, что колебания затухаются. Пусть вышеприведенное неравенство делится на величину $\frac{n}{\log n}$:

$$c_1 < \frac{\pi(n)}{\left(\frac{n}{\log n}\right)} < c_2.$$

Если вычислить значения выражения, стоящего посередине неравенства при $n = 10$, 100, 1000 и 10 000, то получатся следующие значения (приближенно): 0,921, 1,151; 1,188; 1,141, т. е. значения исследуемой функции колеблются около 1. Здесь действительно можно говорить «о прыжках», так как знаменатель всегда увеличивается ненамного, если значение n увеличивается на 1, в то время как числитель увеличивается только в том случае, если наступает очередь нового неделимого числа, но в таком случае он увеличивается не «ненамного», а на 1. Их частное все-таки ограничивается вышепредставленными границами, так как знаменатель всегда растет более медленно, а числитель — все более редко.

Теорема неделимых чисел. Было показано, что значение дроби, приведенной выше, близко к 1. Фактически действительна так называемая *теорема о неделимых числах*, по которой при n, стремящемся к бесконечности, существует граничное значение вышеприведенной дроби, и таким граничным значением является 1. Иными словами: ростом n числитель и знаменатель дроби, представленной выше, будут «все более близки друг к другу», т. е. *функция* $\pi(n)$ *асимптотически равняется выражению* $n/(\log n)$:

$$\pi(n) \sim \frac{n}{\log n}.$$

Итак, с помощью теоремы о неделимых числах возникает возможность исследования функции, рассматриваемой более просто, вместо функции $\pi(n)$ при больших значениях n. Асимптотические равенства имеют особое значение в области теории чисел.

С помощью теоремы о неделимых числах утверждается существенно больше, чем неравенством, представленным в отношении функции $\pi(n)$. Поэтому понятно, что ее доказательство также намного труднее, и для доказательства необходимо применение большого количества вспомогательных средств. Теорема о неделимых числах впервые была доказана Адамар и Ла Валле Пуссен в 1896 г.

ТЕОРЕМА ЧЕБЫШЕВА

Одним из лучших доказательств вышеприведенного неравенства для неделимых чисел является такой способ доказательства, при котором показывается, что между некоторым числом, имеющим значение больше 1 и его двукратным значением расположено «довольно большое количество» неделимых чисел. Проще говоря, между любым числом, имеющим значение больше 1, и его двукратным значением имеется неделимое число.

Настоящая теорема называется *теоремой Чебышева.* Доказательство, выведенное венгерским ученым Палом Эрдёшем, считается особенно удачным и простым. Приведем основную идею доказательства и представим два частных случая.

Доказательство. Биномиальные коэффициенты записываются в виде (см. главу «Комбинаторика»):

$$\binom{n}{k} = \frac{n(n-1)\ldots(n-k+1)}{1.2.\ldots.k}.$$

Это дает возможность доказать теорему. Пусть рассматривается биномиальный коэффициент:

$$\binom{2n}{n} = \frac{2n \cdot (2n-1) \ldots (n+1)}{1 \cdot 2 \cdot \ldots \cdot n}.$$

Данное выражение представляет собой не дробь, а целое число. При сокращении из неделимых сомножителей в знаменателе остаются «немногие». Таким образом, «существенную часть» разложения выражения $\binom{2n}{n}$ на произведение неделимых сомножителей составляют неделимые числа между n и $2n$, что может быть показано путем применения удачно выбранных, но совершенно элементарных оценок. Ниже рассматриваются случаи при $n = 10$ и $n = 100$. В случае $n = 10$ после сокращения получится следующий результат:

$$\binom{20}{10} 4 \cdot 11 \cdot 13 \cdot 17 \cdot 19.$$

Итак, существенно большая часть произведения определяется произведением неделимых чисел, расположенных между 10 и 20, чем неделимых чисел, значение которых меньше чем 10. Вычисление случая при $n = 100$ более трудоемко. Проводя расчеты, получаем следующее выражение:

$$\binom{200}{100} = A \cdot B,$$

где $A = 2^3 \cdot 3 \cdot 5 \cdot 11 \cdot 13^2 \cdot 17 \cdot 37 \cdot 59 \cdot 61$ и
$B =$ произведение неделимых чисел, расположенных между 100 и 200.

Можно вычислить, что A представляет собой число, состоящее примерно из 10 цифр, в то время как в числе AB находится примерно 60 цифр. Следовательно, «существенная часть» произведения и в этом случае определяется произведением неделимых чисел, расположенных между n и $2n$.

СУММА ОБРАТНЫХ ЗНАЧЕНИЙ НЕДЕЛИМЫХ ЧИСЕЛ

Как мы видели, неделимые числа расположены «достаточно редко» среди натуральных чисел. Приведем такое утверждение, которое подчеркивает, что неделимые числа «не настолько редко» расположены. Известно (см. главу «Анализ»), что сумма обратных значений целых чисел как бесконечная прогрессия расходится, что является следствием того, что хотя обратное значение больших чисел является малым, но количество таких значений «достаточно велико» для того, чтобы их сумма стала расходящейся. Наоборот, о сумме обратных значений квадратных чисел может быть сказано, что она сходится (меньше чем 2). Причина в том, что количество квадратных чисел «слишком мало» для того, чтобы бесконечная прогрессия, образованная из их обратных значений, стала расходящейся.

Обозначим через $f(n)$ сумму обратных значений неделимых чисел, которые не больше чем n. Расходимость функции $f(n)$, точнее, утверждение, по которому $\lim\limits_{n\to\infty} [f(n)] = \infty$, указывает на то, что существует большое количество неделимых чисел. Известно множество приемов этого доказательства. Здесь не приводится ни один из этих приемов, только делается замечание в отношении двух доказательств. Самое элементарное доказательство теоремы принадлежит венгерскому ученому Палу Эрдёшу, и хотя в нем нет большого количества вспомогательных средств, оно включает в себя много таких понятий, представление которых потребовало бы продолжительных дискуссий. Другое доказательство приведено Эйлером; в нем необходимо применение большого числа вспомогательных средств, но на основе этого доказательства можно показать правильность асимптотического равенства:

$$f(n) \sim \log [\log(n)] .$$

Настоящий результат может быть получен с помощью исследования определенного бесконечного произведения. Из обобщения рассматриваемого бесконечного произведения получится так называемая ζ-функция (читается: дзета-функция), играющая важную роль в области теории чисел. На основе исследования расположения корней данной функции могут быть сделаны очень важные выводы. Но расположение этих корней неизвестно, хотя многими известными математиками были сделаны попытки доказательства так называемой *догадки Римана*, относящейся к данному вопросу.

Мало известно и о точном расположении неделимых чисел. В то же время и в этой области доказана правильность одного асимптотического равенства, по которому:

$$p_n \sim n \cdot \log n,$$

где p_n обозначает n-ое неделимое число.

Неделимые числа-близнецы. Нерегулярное поведение неделимых чисел указывается существованием так называемых *неделимых чисел-близнецов*. Поскольку единственным четным неделимым числом является 2, то разность двух неделимых чисел в единственном случае (2 и 3) может составлять 1. Но чаще встречаются такие случаи, когда разность двух неделимых чисел — 2. Такими являются, например, 3 и 5; 5 и 7; или 101 и 103. Такие пары неделимых чисел называются неделимыми числами-близнецами. Вопрос о том, существует ли бесконечное количество неделимых чисел-близнецов, остается открытым. О таких числах нам известно только то, что их количество не может быть слишком большим, так как сумма их обратных значений сходится.

НЕДЕЛИМЫЕ ЧИСЛА АРИФМЕТИЧЕСКИХ ПРОГРЕССИЙ

Существует единственное четное неделимое число — 2. Все остальные неделимые числа нечетны. Нечетные неделимые числа делятся на два класса. Если поделить нечетное число на 4, в остатке результата появляется или 1, или 3. Соответственно, все неделимые нечетные числа могут быть педставлены или в виде $4k+1$, или — $4k+3$, где k — неотрицательное целое число. Возникает вопрос: количество неделимых чисел такого представления является ли ограниченным или бесконечным?

Легко показать, что существует бесконечное количество неделимых чисел вида $4k+3$. Доказательство аналогично доказательству существования бесконечного количества неделимых чисел. Следует учесть, что произведение двух чисел вида $4k+1$ имеет такой же вид. Действительно, $(4k+1)(4n+1)=4(4kn+n+k)+1$, в соответствии с данным утверждением. А если бы $p_1, p_2, \ldots, p_r$ представляли собой все неделимые числа вида $4k+3$, то пусть рассматривается число $4p_1p_2\ldots p_r-1$. Данное число представляется в виде $4k+3$, поэтому оно не может быть произведением неделимых чисел, имеющих вид $4k+1$. Следовательно, оно имеет неделимый делитель вида $4k+3$, отличающийся, очевидно, от заданных неделимых чисел, из чего уже следует правильность утверждения.

Труднее доказать, что существует бесконечное количество неделимых чисел вида $4k+1$. Для доказательства такого утверждения используется свойство, по которому каждый неделимый делитель некоторого числа вида n^2+1 имеет вид $4k+1$. (Настоящее свойство еще будет применено при представлении конгруэнций второго порядка.) Доказательство проводится подобно предыдущему, с той лишь разницей, что для заданных неделимых чисел $p_1, p_2\ldots, p_r$ вида $4k+1$ следует рассматривать выражение $(2p_1p_2\ldots p_r)^2+1$.

Теорема Дирихле. Может быть исследована проблема: существует ли бесконечное количество неделимых чисел вида $ak+b$, где a и b — заданные натуральные числа, а k принимает все значения неотрицательных целых чисел. Поскольку вышеприведенные числа представляют собой арифметическую прогрессию, вопрос заключается в том, существует ли бесконечное количество неделимых чисел в некоторых арифметических прогрессиях. Рассмотрим, при каких значениях a и b не существует бесконечное количество неделимых чисел в исследуемой арифметической прогрессии.

Пусть $d=(a, b)$. Если $d\neq 1$, то в арифметической прогрессии, приведенной выше, имеется, по крайней мере, одно неделимое число. Действительно, из-за условия каждое число вида $ak+b$ делится на d, и поскольку все они различные, по крайней мере одно из них может равняться числу d. А если число d присутствует в вышеприведенной арифметической прогрессии и представляет собой неделимое число, то в данной прогрессии имеется одно неделимое число, иначе — ни одного нет. В остальных случаях можно показать, что в арифметической прогрессии имеется бесконечное количество неделимых чисел, т. е. *в некоторой арифметической прогрессии, состоящей из натуральных чисел, количество неделимых чи-*

сел 0, 1 *или бесконечно*. В последнем случае можно показать, что количество неделимых чисел, расположенных в арифметической прогрессии, зависит — асимптотически — только от разности арифметической прогрессии. Точнее, если $(a, b) = 1$ и $\pi_{a,b}(n)$ обозначает количество неделимых чисел, значение которых меньше n, и имеющих вид $ak + b$, то действительно соотношение: $\pi_{a,b}(n) \sim \dfrac{n}{\varphi(a)\cdot\log n}$, где φ — функция Эйлера (см. главу «Функции в теории чисел»). Настоящая теорема называется *теоремой Дирихле*.

НЕДЕЛИМЫЕ ЧИСЛА РАЗЛИЧНЫХ ТИПОВ

Результат, приведенный в предыдущей части, может быть сформулирован так: если $(a, b) = 1$, то значения многочлена $ax + b$ будут неделимыми числами при замене бесконечного количества целых чисел. Приведенный многочлен является многочленом первой степени. Как будет обстоять дело, если рассматривать многочлен не первой, а более высокой степени? И здесь легко определить такие многочлены, в отношении которых утверждение видимо будет неправильно. Если следует найти многочлен, в отношении которого утверждение является правильным, то целесообразно исследовать самый простой многочлен не первой степени. Таким многочленом является (при исключении тривиального случая, упомянутого выше) многочлен $x^2 + 1$. Уже в этом простейшем случае нам ничего неизвестно; вопрос о том, существует ли бесконечное количество неделимых чисел вида $k^2 + 1$, остается открытым.

Неделимые числа типа Ферма. В различных областях математики рассматривались неделимые числа и других типов. Из них наиболее важными считаются неделимые числа двух типов: типа Ферма и типа Мэрзенна. При применении некоторых элементарных алгебраических преобразований число вида $2^n + 1$ может быть неделимым только тогда, если число n само является степенью 2. Неделимые числа вида $2^{2^k} + 1$ называются *неделимыми числами типа Ферма*; они важны в теории конструирования правильных многоугольников.

Легко убедиться в том, что при $k = 0, 1, 2, 3$ и 4 получаются неделимые числа, а при $k = 5$ — нет: $641 | 2^{2^5} + 1$.

Возникает вопрос, существует ли еще неделимое число типа Ферма. Поскольку каждое неделимое число типа Ферма имеет вид $k^2 + 1$, то безнадежно принимать решение о том, существует ли бесконечное количество неделимых чисел типа Ферма.

Неделимые числа типа Мэрзенна. Другим интересным типом неделимых чисел являются неделимые числа вида $2^n - 1$. Используя простые алгебраические преобразования, также можно убедиться, что число такого вида может быть неделимым только тогда, если $n = p$ также является неделимым числом. Неделимые числа вида $2^p - 1$ называются *неделимыми числами типа Мэрзенна*. Такие неделимые числа представляют собой интерес и потому, что «самые большие известные неделимые числа» имели такой вид. Долгое время рекордом считалось число $2^{127} - 1$.

Затем с появлением электронно-вычислительных машин удалось найти более крупные неделимые числа такого типа. Три из них приводятся здесь: $2^{2281}-1$ (которое также считался рекордом долгое время), $2^{3217}-1$ и $2^{4423}-1$. Последнее число считается самым большим неделимым числом, известным в настоящее время.

Наконец, еще одно интересное свойство неделимых чисел типа Мэрзенна. Рассмотрим следующую последовательность: 2, $2^2-1=3$, $2^3-1=7$, $2^7-1=127$, $2^{127}-1, \ldots$ Каждый элемент данной последовательности является неделимым числом, и если p — один из ее элементов, то следующий ее элемент определяется согласно 2^p-1, т. е., за исключением первого ее элемента, каждый элемент является неделимым числом типа Мэрзенна. Будет ли неделимым числом следующий ее элемент, будет ли каждый элемент данной последовательности чисел неделимым числом, — пока неизвестно.

ФУНКЦИИ В ТЕОРИИ ЧИСЕЛ

ФУНКЦИИ В ТЕОРИИ ЧИСЕЛ И ИХ НЕКОТОРЫЕ СПЕЦИАЛЬНЫЕ КЛАССЫ

В области неделимых чисел было представлено несколько функций теории чисел. Такой функцией являлось, например, количество неделимых чисел, значение которых не больше n; или сумма обратных значений неделимых чисел, значения которых не больше n. В обоих случаях существенным считалось свойство, по которому данные функции существуют только в области натуральных чисел, но в их значения входили не только натуральные числа. В общем случае:

Функциями теории чисел называются функции, областью существования которых является совокупность натуральных чисел, а их значениями — действительные числа.

Отметим, что во многих случаях допускается более обширная область значений.

Мультипликативные функции теории чисел. Среди функций в области теории чисел весьма важными считаются так называемые *мультипликативные функции теории чисел.*

Функция теории чисел $f(n)$ называется мультипликативной, если при любой паре чисел a и b, являющихся относительно неделимыми числами, выполняется условие $f(ab)=f(a)f(b)$, и функция $f(n)$ не имеет значение 0 тождественно. В случае, когда равенство $f(ab)=f(a)f(b)$ выполняется при любых целых числах a и b, то функция $f(n)$ теории чисел называется абсолютно мультипликативной.

Абсолютно мультипликативной является, например, функция $f(n)=n^2$. Ясно, что каждая абсолютно мультикативная функция является одновременно и мультипликативной, но бывают не абсолютно мультипликативные функции, которые являются мультипликативными. (Ниже будут представлены примеры мультипликативных функций теории чисел.)

Покажем, *что если $f(n)$ представляет собой мультипликативную функцию теории чисел, то $f(1) = 1$.* Если $f(n)$ не равняется 0 тождественно, то существует такое натуральное число a, в отношении которого выполняется условие $f(a) \neq 0$. Если $(a, 1) = 1$, то получится соотношение $f(a) = f(a \cdot 1) = f(a)f(1)$, и если сократить на значение $f(a)$, отличающееся от 0, то получится желаемое равенство.

При мультипликативных функциях теории чисел нельзя задавать произвольно значения функции на всех местах существования. Если, например, $f(2) = 5$ и $f(3) = 2$, то из-за мультипликативности получится значение $f(6) = 10$. Учитывая данное замечание, появляется возможность вычислить значения некоторой мультипликативной функции теории чисел и в том случае, когда эти значения известны лишь в некоторых местах существования функции. Точнее:

Если значения некоторой мультипликативной функции теории чисел известны при степенях неделимых чисел, то ее значения могут быть вычислены на всех местах существования, где аргумент имеет целое число. Именно, если $n = p_1^{\alpha_1} \cdot p_2^{\alpha_2} \ldots p_r^{\alpha_r}$ является канонической формой n, то имеет место соотношение:

$$f(n) = f(p_1^{\alpha_1}) \cdot f(p_2^{\alpha_2}) \cdot \ldots \cdot f(p_r^{\alpha_r}),$$

что является естественным следствием определения мультипликативности, если учесть, что степени различных неделимых чисел являются попарно относительными неделимыми числами в отношении друг к другу.

Аддитивные и абсолютно аддитивные функции теории чисел. Подобно определению мультипликативных функций теории чисел можно приводить определения аддитивных и абсолютно аддитивных функций теории чисел. Но эти функции являются менее важными, чем предыдущие.

Функция теории чисел $f(n)$ называется аддитивной, если при $(a, b) = 1$ выполняется соотношение $f(ab) = f(a) + f(b)$ и абсолютно аддитивной, если данное условие выполняется без какого-либо ограничения.

ОСНОВНЫЕ МУЛЬТИПЛИКАТИВНЫЕ ФУНКЦИИ ТЕОРИИ ЧИСЕЛ

Термин мультипликативной функции теории чисел был введен, так как было замечено, что различные важные функции в области теории чисел имеют свойство мультипликативности. Ниже приводятся сведения об основных таких функциях.

φ-ФУНКЦИЯ ЭЙЛЕРА

Пусть n представляет собой натуральное число и пусть обозначается символом $\varphi(n)$ количество относительных неделимых чисел в отношении к значению n и имеющих значение не больше, чем n.

Например, $\varphi(1) = 1$, так как рассматривается исключительно число 1 и $(1, 1) = 1$; $\varphi(2) = 1$, так как рассматриваются лишь числа 1 и 2 и $(1, 2) = 1$, $(2, 2) \neq 1$; $\varphi(4) = 2$, так как в отношении рассматриваемых чисел 1, 2, 3, 4 имеет место соотношение $(1, 4) = (3, 4) = 1$; но 2 и 4 не являются относительными неделимыми числами в отношении к 4; $\varphi(12) = 4$, так как от 1 до 12 имеются четыре относительных неделимых числа в отношении к 12: 1, 5, 7 и 11.

Ниже будет показано, что функция Эйлера является мультипликативной. Пусть будет $(a, b) = 1$ и пусть записываются числа от 1 до ab следующим образом:

$$\begin{array}{rrcr}
1, & 2, & \ldots & b \\
b+1, & b+2, & \ldots & 2b \\
\vdots & & & \\
b(a-2)+1, & b(a-2)+2, & \ldots & (a-1)b \\
b(a-1)+1, & b(a-1)+2, & \ldots & ab.
\end{array}$$

Среди этих чисел следует найти относительно неделимые числа, являющиеся относительно неделимыми в отношении к ab. Из основной теоремы теории чисел следует, что некоторое число c является относительно неделимым к числу ab тогда и только тогда, если оно является относительно неделимымым в отношении как к числу a, так и к числу b.

Вначале найдем те числа среди вышеприведенных, которые являются относительно неделимыми по отношению к числу b, а затем среди них выявим те, которые являются относительно неделимыми по отношению к числу a. Из свойств НБОД следует, что $(c, b) = (c + b, b)$. Следовательно, если в таблице, представленной выше, некоторое число является относительно неделимым в отношении к числу b, то каждое такое число будет относительно неделимым по отношению к числу b, которое расположено в том же столбце. Поэтому все числа, относительно неделимые по отношению к числу b, могут быть получены из вышеприведенной таблицы следующим обазом: в первой строке рассматриваются числа, относительно неделимые по отношению к числу b, и если c — такое число, то возьмем его и все числа, стоящие под ним; получится следующее:

$$c,\ b + c,\ 2b + c, \ldots, (a - 1)b + c.$$

При обсуждении проблем конгруэнций будет показано, что при числах a и b, относительно неделимых по отношению друг к другу, среди этих чисел имеются столько же чисел относительно неделимых по отношению к числу a, как и среди чисел $1, 2, \ldots, a$, т. е. всего $\varphi(a)$. Поскольку в первой строке по определению имеются всего $\varphi(b)$ чисел, относительно неделимых по отношению к числу b, то числа, относительно неделимые по отношению к числу ab, расположены в столбцах, количество которых составляет $\varphi(b)$, и в каждом столбце имеются всего $\varphi(a)$ таких чисел, поэтому их общее число составляет $\varphi(a)\varphi(b)$. Таким образом доказана мультипликативность функции Эйлера. Но необходимо заметить, что функция Эйлера не является абсолютно мультипликативной: например, $\varphi(2) = 1$, но $\varphi(4) = 2 \neq 1 \cdot 1 = \varphi(2)\varphi(2)$.

Чтобы определить значение функции $\varphi(n)$ при любом натуральном числе n, необходимо только задать значение $\varphi(p^\alpha)$ (p — любое неделимое число; α — натуральное число). От 1 до p^α имеются всего p^α целых чисел. Если следует определить, сколько из них является числом, относительно неделимым по отношению к числу p^α, то из их количества, т. е. из p^α следует вычитать количество тех чисел, которые не являются относительно неделимыми по отношению к p^α. Поскольку любой делитель p^α может быть только некоторая степень p, то числами, не являющимися относительно к p^α неделимыми, могут быть те и только те числа, которые делятся на p. До p^α такие числа определяются согласно:

$$1 \cdot p,\ 2 \cdot p, \ldots, p^{\alpha-1} \cdot p \ (= p^\alpha)\,.$$

Количество этих чисел составляет $p^{\alpha-1}$ и, таким образом, $\varphi(p^\alpha) = p^\alpha - p^{\alpha-1}$. Целесообразно вынести из данной разности значение p^α, после такого преобразования имеет место соотношение: $\varphi(p^\alpha) = p^\alpha \left(1 - \dfrac{1}{p}\right)$. При любом целом числе n, большем 1, значение функции $\varphi(n)$ определяется таким образом:

Если $n = p_1^{\alpha_1} \cdot p_2^{\alpha_2} \cdot \ldots \cdot p_r^{\alpha_r}$, то

$$\varphi(n) = n\left(1 - \frac{1}{p_1}\right)\left(1 - \frac{1}{p_2}\right) \cdot \ldots \cdot \left(1 - \frac{1}{p_r}\right).$$

μ-ФУНКЦИЯ МЁБИУСА

Данная функция, важная в теории чисел, определяется соотношением:

$$\mu(n) = \begin{cases} 1, & \text{если } n = 1 \\ (-1)^r, & \text{если } n = p_1 p_2 \ldots p_r \\ 0 & \text{в любом другом случае.} \end{cases}$$

Из определения вытекает, что условие $\mu(n) = 0$ выполняется тогда и только тогда, если число n делимо на некоторое квадратное число. Легко убедиться в том, что функция Мёбиуса действительно является мультипликативной. Именно поэтому данная функция могла бы быть определена как *мультипликативная функция, принимающая значения* -1, *где аргумент является неделимым числом, и* 0, *где аргумент является степенью некоторого неделимого числа*. Функция Мёбиуса не является абсолютно мультипликативной.

КОЛИЧЕСТВО И СУММА ДЕЛИТЕЛЕЙ

Число (положительных) делителей натурального числа n обозначается через $d(n)$, а сумма делителей — через $S(n)$. Обе функции теории чисел являются мультипликативными, но не являются абсолютно мультипликативными. При известном разложении числа n на произведение неделимых чисел, значения обеих функций могут быть записаны в виде:

Если $n = p_1^{\alpha_1} \cdot p_2^{\alpha_2} \cdot \ldots \cdot p_r^{\alpha_r}$, то $d(n) = (\alpha_1 + 1)(\alpha_2 + 1) \ldots (\alpha_r + 1)$ и

$$S(n) = \frac{p_1^{\alpha_1+1} - 1}{p_1 - 1} \cdot \frac{p_2^{\alpha_2+1} - 1}{p_2 - 1} \cdot \ldots \cdot \frac{p_r^{\alpha_r+1} - 1}{p_r - 1}.$$

КОНГРУЭНЦИИ

ПОНЯТИЕ КОНГРУЭНЦИИ

В ходе обсуждения предыдущих частей часто встречались такие случаи, когда следовало учитывать лишь то, что получится остатком в результате деления некоторого целого числа на другое целое число (например, в евклидовом алгоритме). Подобное положение возникает при решении задач по делимости или задач пробы по 9 и 11, которые будут рассмотрены после обсуждения темы конгруэнции. В таких случаях целесообразно не различать числа, которые при делении на некоторое фиксированное число дают одинаковый остаток. (В таком случае эти числа называются *конгруэнтными* по отношению к делителю.) Если целые числа a и b дают при делении на положительное целое число m остаток r $(0 \leqq r < m)$, то эти числа могут быть записаны в виде:

$$a = cm + r \quad \text{и} \quad b = dm + r.$$

Тогда $a - b = (cm + r) - (dm + r) = cm - dm = (c - d)m$, то есть $m|(a - b)$. Данное соотношение действительно и в его обратном виде, т. к. u и v представляют собой такие целые числа, в отношении которых имеет место соотношение $m|(u - v)$. Кроме того, пусть выполняются условия $u = pm + s$ и $v = qm + t$, где p, q, s и t — целые числа, среди них как s, так и t меньше m и неотрицательны. В соответствии с условием выражение $u - v = (pm + s) - (qm + t) = (p - q)m + (s - t)$ является делимым на число m. Поскольку выражение $(p - q)m$ очевидно делимо на m, то выражение $s - t$ также представляет собой число, делимое на m. Учитывая, что как s, так и t положительны и меньше m, то разность $s - t$ обязательно будет больше, чем $-m$ и меньше, чем $+m$. А между данными двумя числами имеется единственное число, делимое на m, а именно: 0. Итак, $s - t = 0$, т. е. $s = t$, следовательно, при делении двух заданных чисел на m у обоих получится одинаковый остаток. Такая формулировка является более удобной с точки зрения определения, формулируемого так:

Целые числа a и b называются конгруэнтными по модулю m (с точки зрения модуля по m), если выполняется условие $m|(a-b)$ (m — целое число).

Обозначением соотношения, по которому a и b — конгруэнтны по модулю m, является следующее:

$$a \equiv b \pmod m.$$

Невыполнение условия конгруэнтности обозначается в виде:

$$a \not\equiv b \pmod m.$$

При обсуждении проблематики конгруэнтности *модуль* m считается фиксированным. Поскольку при исследованиях по делимости знак чисел не рассматривается, то достаточно ограничиться сведениями по неотрицательным модулям. Если $m = 0$, то вражение $a \equiv b \pmod 0$ означает, что $0|(a-b)$, что выполнимо только в случае $a = b$, т. е. *конгруэнтность по модулю* 0 *означает равенство и, таким образом, равенство можно считать специальным случаем конгруэнтности.*

Если $m = 1$, то выражение $a \equiv b \pmod 1$ означает $1|(a-b)$, а данное соотношение всегда правильно, т. е. *любые два целых числа являются конгруэнтными по модулю* 1. Рассмотрим случай $m = 2$. Выражение $a \equiv b \pmod 2$ означает, что $2|(a-b)$, т. е. $a-b$ является четным числом. Данное соотношение действительно только тогда, если a и b или оба являются четными числами, или оба — нечетными. Два числа называются числами одинаковой *паритетности* или одинаковой *четности*. Итак, *два числа являются конгруэнтными по модулю* 2, *если их паритетность совпадает.*

КЛАССЫ ОСТАТКОВ

По исходному предположению, два числа считаются тождественными (с определенной точки зрения) при фиксированном значении модуля m, если в результате деления их на m получаются одинаковые остатки. Но это не значит, что эти числа станут тождественными, раз они принадлежат к одинаковому множеству в отношении рассматриваемого свойства; эти числа *принадлежат к одному классу.*

Следовательно, к одному классу принадлежат те числа, у которых в результате деления на m обнаруживается один и тот же остаток.

Например, по модулю 5 получаются следующие классы, так называемые *классы остатков:*

$$\begin{array}{lrrrrl}
\ldots, & -10, & -5, & 0, & 5, & 10, \ldots \\
\ldots, & -9, & -4, & 1, & 6, & 11, \ldots \\
\ldots, & -8, & -3, & 2, & 7, & 12, \ldots \\
\ldots, & -7, & -2, & 3, & 8, & 13, \ldots \\
\ldots, & -6, & -1, & 4, & 9, & 14, \ldots
\end{array}$$

Разделение чисел на классы означает следующее:

1. Каждое число принадлежит к некоторому классу.
2. Каждое число принадлежит только к одному классу.

По определению числа a и b принадлежат к одному классу только в том случае, если $a \equiv b \pmod{m}$. Конгрэнтности такого характера представляют собой связь, существующую между двумя числами (по-латински *реляцию* — *соотношение).*

Элементарные свойства конгруэнтности. Разделением чисел на классы выражаются следующие три свойства конгруэнтности:

1. $a \equiv a \pmod{m}$ — конгруэнтность является рефлексивной.
2. Если $a \equiv b \pmod{m}$, то $b \equiv a \pmod{m}$ — когруэнтность является симметричной.
3. Если $a \equiv b \pmod{m}$ и $b \equiv c \pmod{m}$, то $a \equiv c \pmod{m}$ — конгруэнтность является транзитивной.

Если некоторое соотношение является рефлексивным, симметричным и транзитивным, то оно называется *соотношением эквивалентности.* Соотношением эквивалентности является, например, кроме конгруэнтностей, равенство между числами, или равенство и подобие между плоскими фигурами, и параллельность между прямыми пространства.

При фиксированном значении m конгруэнтность характеризуется следующим свойствами:

Если $a \equiv b \pmod{m}$, то $a + c \equiv b + c \pmod{m}$, так как $(a + c) - (b + c) = a - b$ делимо на m.

Если $a \equiv b \pmod{m}$ и $c \equiv d \pmod{m}$, то $(a + c) \equiv b + d \pmod{m}$, так как из предыдущего свойства следует правильность соотношений $a + c \equiv b + c \pmod{m}$ и $b + c \equiv b + d \pmod{m}$, а затем из свойства транзитивности следует само утверждение.

Если $a \equiv b \pmod{m}$, то $ac \equiv bc \pmod{m}$, так как, сложив конгруэнтность c раз (на основе предыдущего свойства) получится опять правильная конгруэнтность.

Если $a \equiv b \pmod{m}$ и $c \equiv d \pmod{m}$, то $ac \equiv bd \pmod{m}$.

Если $a \equiv b \pmod{m}$ и n представляет собой натуральное число, то $a^n \equiv b^n \pmod{m}$.

Оба последних свойства могут быть доказаны аналогично предыдущим доказательствам.

Итак рассматривались следующие свойства:

Если к правильной конгруэнтности прибавить некоторое число, умножать ее на некоторое число, или возвести ее в степень, то опять получится правильная конгруэнтность.

Суммируя или умножая правильные конгруэнтности, получим правильную конгруэнтность.

Приведем еще и такой легко доказуемый результат: если $a \equiv b \pmod{m}$ то $(a, m) \equiv (b, m)$.

ПРАВИЛА ДЕЛИМОСТИ

Если исследуется вопрос, является ли целое число N делимым на некоторое натуральное число b, то необходимо отметить, будет ли в результате деления числа на число b остаток 0 или нет, т. е. имеет ли место конгруэнтность $N \equiv 0 \pmod{b}$. (В общем случае N представляет собой натуральное число.) При рассуждении мы не ограничиваемся случаем, когда остаток является 0; посмотрим в общем случае, что собой представляет остаток деления. Пусть с этой целью записывается число N в виде числа в десятичной системе чисел:

$$N = a_n \cdot 10^n + a_{n-1} \cdot 10^{n-1} + \ldots + a_2 \cdot 10^2 + a_1 \cdot 10 + a_0,$$

где $a_n, a_{n-1}, \ldots, a_2, a_1, a_0$ — отдельные цифры в числе N, следовательно, они представляют собой неотрицательные целые числа меньше 10. (Например, $53\,039 = 5 \cdot 10^4 + 3 \cdot 10^3 + 0 \cdot 10^2 + 3 \cdot 10 + 9$.)

а) Делимость на 2, 4 и 8.

Рассмотрим сначала делимость на 2. Поскольку $10 \equiv 0 \pmod{2}$, то, учитывая свойства конгруэнтности, $N = a_0 \pmod{2}$. т. е. некоторое натуральное число является четным (делимым на 2) тогда и только тогда, если последняя его цифра является четным числом.

Конгруэнтности $N \equiv a_1 \cdot 10 + a_0 \cdot \pmod{4}$ и $N \equiv a_2 \cdot 10^2 + a_1 \cdot 10 + a_0 \pmod{8}$ при рассмотрении вопроса о делимости на 4 и 8 получатся подобно тому, как и в предыдущем случае. Поэтому

некоторое натуральное число является делимым на 4 и на 8, если число, состоящее из последних двух и трех цифр исходного числа, соответственно, является делимым на 4 и на 8.

б) При исследовании вопроса о делимости на 5, 25 и 125 получатся аналогичные результаты.

в) Делимость на 3 и на 9.

Поскольку из конгруэнтности $a \equiv b \pmod{9}$ следует конгруэнтность $a \equiv b \pmod{3}$, то оказывается достаточным посмотреть остаток деления на 9. Учитывая, что $10 - 1 = 9$, $- 10 \equiv 1 \pmod{9}$. На основе свойства конгруэнтности в отношении возведения конгруэнтности в квадрат можно сделать вывод, что при любом натуральном числе k действительно соотношение $10^k = 1 \pmod{9}$. Учитывая данное соотношение, получится следующее:

$$N \equiv a_n + a_{n-1} + \ldots + a_2 + a_1 + a_0 \pmod{9}.$$

Иными словами, *остаток деления любого натурального числа на* 9 (*на* 3) *совпадает с остатком деления суммы знаков числа на* 9 (*на* 3). Т. е. некоторое натуральное число является делимыми на 9 (на 3) тогда и только тогда, если сумма его знаков является делимым на 9 (на 3).

Например, сумма знаков числа 45 726 будет: $4 + 5 + 7 + 2 + 6 = 24$, следовательно, при делении числа 45 726 на 9 и на 3 остаток будет такой же, как и при делении числа 24. Далее, оказывается, сумма знаков 24: $2 + 4 = 6$, откуда следует, что при делении 45 726 на 9 остаток будет 6 (итак, оно неделимо на 9) и оно является делимым на 3 (так как число 6 делимо на 3).

г) Делимость на 11.

Из конгруэнтности $10 \equiv -1 \pmod{11}$ следует, что при любом натуральном числе k имеет место конгруэнтность $10^k \equiv (-1)^k \pmod{11}$. Исходя из этого:

$$N \equiv a_0 - a_1 + a_2 - \ldots + (-1)^{n-1}a_{n-1} + (-1)^n a_n \pmod{11}.$$

Итак, если брать знаки некоторого числа в обратном порядке с чередованием знака, то остаток деления полученного таким образом числа на 11 совпадает с остаком деления на 11 исходного числа. При предыдущем числе 45 726 данная сумма составляет: $6 - 2 + 7 - 5 + 5 = 10$, поэтому при делении 45 726 на 11 остаток будет 10.

д) Пробы на 9 и 11.

Делимость применяется и для того, чтобы с ее помощью контролировать правильность операций. Рассмотрим проблему контроля произведения. Предположим, что при умножении чисел A и B получен результат C. Пусть, далее, обозначаются через a, b и c остатки деления этих чисел на модуль m. С помощью соотношения произведения конгруэнтностей, из этого следует конгруэнтность $ab \equiv c \pmod{m}$. Данный результат формулируется так:

произведение остатков сомножителей является конгруэнтным остатку произведения.

Если операция умножения осуществлена правильно, то данное соотношение должно иметь место и, наоборот, если данное соотношение не выполняется, то в ходе операции умножения допущена ошибка. Если выбирать модули на 9 и 11, то они называются пробами на 9 и 11. При осуществлении таких проб правильность результата станет более достоверной, хотя и при выполнении ошибочной операции умножения может случиться, что по обеим пробам утверждается правильность сделанной операции. Эти две пробы целесообразно проводить потому, что при обоих модулях легко определить остаток, и поскольку при определении фигурирует каждая цифра, то при *любой* описке появляется ошибка.

Примеры:

5362 · 4571	435 · 322	354 · 217
37534	880	2378
26810	880	354
21448	1105	708
24509702	120180	76718

Во всех трех произведениях приведены пробы на 9. В первом случае остаток составляет $5 + 3 + 6 + 2 = 16$, $4 + 5 + 7 + 1 = 17$ и $2 + 4 + 5 + 0 + 9 + 7 + 0 + 2 = 29$; остатки этих остатков составляют: $1 + 6 = 7$, $1 + 7 = 8$ и $2 + 9 = 11$ (а остаток последнего: $1 + 1 = 2$). Если перемножить остатки друг на друга, результат имеет вид: $7 \cdot 8 = 56$ и остаток от числа 56 ($5 + 6 = 11$, $1 + 1 = 2$) действительно составляет 2. Итак, в первом примере с помощью пробы на 9 ошибка не указывается. Если подсчитать остатки во втором примере, то получаются 3, 7 и 3, соответственно. Поскольку остатком числа $3 \cdot 7 = 21$ является 3, то пробой на 9 и в этом примере ошибка не указывается. В третьем примере остатки будут 3, 1 и 2. Учитывая, что остатком числа $3 \cdot 1 = 3$ *не* является 2, третье умножение наверняка ошибочно. Таким образом, с помощью пробы на 11 следует контролировать правильность лишь первых двух примеров. В первом примере остатки имеют значения: $2 - 6, + 3 - 5 = -6$ (т. е. 5), $1 - 7 + 5 - 4 = -5$ (т. е. 6) и $2 - 0 + 7 - 9 + 0 - 5 + 4 - 2 = -3$ (т. е. 8). Действительно, остатком числа $6 \cdot 5 = 30$ является 8. Во втором примере остатки: $5 - 3 + 4 = 6$, $2 - 2 + 3 = 3$ и $0 + 1 - 0 + 2 - 1 = -6$ (т. е. 5). Но с другой стороны остатком числа $6 \cdot 3 = 18$ является $8 - 1 = 7$, а *не* 5, следовательно, данное умножение также ошибочно (хотя ошибка еще не была выявлена с помощью пробы на 9). Если размеры ошибки достигают значения 99 ($= 9 \cdot 11$), то такая ошибка не может быть выявлена. Но данные две пробы применяются преимущественно, и если применить обе пробы одновременно, ошибка выявляется в 99% случаев.

е) Делимость на 7, 11 и 13.

Когда рассматривается делимость на вышеприведенные числа, то задача всегда может быть сведена к задаче определения делимости числа, состоящего в крайнем случае из трех знаков. Это обосновывается тем, что $7 \cdot 11 \cdot 13 = 1001$. Подобно тому, как было представлено исследование делимости чисел на 11, можно показать, что, например, при делении числа 45 412 395 на 7, 11 или на 13 будет такой же остаток, как и при делении числа $395 - 412 + 45 = 28$ на эти же числа. Таким образом, данное число при делении на 7 оказывается делимым на 7, при делении на 11 остаток будет 6, а при делении на 13 остаток — 2. Ниже покажем (как следствие теоремы Эйлера), что аналогичное вышеприведенным правило делимости может быть выявлено для всех относительных неделимых чисел, которые являются неделимыми относительно 10.

ПОЛНЫЕ И СОКРАЩЕННЫЕ СИСТЕМЫ ОСТАТКОВ

Было показано, что результаты операций, проводимых на конгруэнтных числах, также являются конгруэнтными (при сложении, при вычитании, при умножении, возведении в степень). Именно этим объясняется тот факт, что если числа рассматриваются по модулю m, достаточно рассматривать лишь одно число среди конгруэнтных между собой чисел.

Полная система остатков. Система целых чисел, в которой по модулю m каждому числу можно найти единственное конгруэнтное ему число, называется полной системой остатков (по модулю m).

Полную систему остатков представляют собой, например, все различные остатки, появляющиеся при делении на m, т. е. числа $0, 1, \ldots, m - 1$

Итак, некоторые числа тогда и только тогда могут представить собой полную систему остатков по модулю m, если они в некотором смысле являются конгруэнтными вышеприведенным числам. Так как при делении на m двух чисел в результатах будут различные остатки лишь в том случае, если они *неконгруэнты* по модулю m, утверждение, представленное выше, может формулироваться так:

Некоторые (целые) числа образуют полную систему остатков тогда и только тогда, по модулю m, если их количество составляет m и они являются попарно неконгруэнтными по модулю m.

В следующей части приводится способ получения новой полной системы остатков из другой полной системы остатков. Впрочем, данный результат уже применялся при доказательстве мультипликативности функции Эйлера.

Если $r_1, r_2 \ldots, r_m$ представляют собой полную систему остатков по модулю m и при этом $(a, m) = 1$, то при любом целом числе b выражение

$$ar_1 + b, ar_2 + b, \ldots, ar_m + b$$

также представляет собой полную систему остатков по модулю m.

Поскольку здесь получены также m чисел, достаточно показать, что они являются попарно неконгруэнтными.

Доказательство. Используется косвенный метод доказательства. Предположим, что два из полученных чисел являются конгруэнтными, значит, и среди исходных чисел имеются два, являющиеся конгруэнтными. Предположим, например, что имеет место конгруэнтность $ar_1 + b \equiv ar_2 + b \pmod{m}$, из чего следует конгруэнтность $ar_1 \equiv ar_2 \pmod{m}$. Если бы можно было делить на a, то получилось бы, что $r_1 \equiv r_2 \pmod{m}$, что представляет собой противоречие. Но в общем случае нельзя осуществить деление, так как хотя и выполняется условие $1 \cdot m \equiv 2 \cdot m \pmod{m}$, но конгруэнтность $1 \equiv 2 \pmod{m}$, получаемая после деления на m, не выполняется. Покажем правильность следующего утверждения:

Если $(a, m) = 1$, то из конгруэнтности $ar \equiv as \pmod{m}$ следует конгруэнтность $r \equiv s \pmod{m}$.

Учитывая, что условие $(a, m) = 1$ выполняется при предыдущей теореме, доказательство настоящей теоремы представляет собой одновременно и конец доказательства предыдущей теоремы. Рассмотрим доказательство настоящей теоремы.

Конгруэнтность $ar \equiv as \pmod{m}$ означает выполнение условия $m|(ar - as)$ или, иначе, $m|ar - s)$.

> Поскольку $(m, a) = 1$, то на основе применения разложения чисел на произведение неделимых сомножителей — $m|(r - s)$, т. е. $r \equiv s \pmod{m}$, в соответствии с утверждением.

С точки зрения делимости настоящий результат подобен той теореме, по которой, если неделимое число является делителем некоторого произведения, то оно является и делителем одного из сомножителей данного произведения. С помощью приема, примененного при доказательстве упомянутой теоремы, и здесь можно достичь результата, из чего можно сделать последующие выводы в отношении конгруэнтности.

Предположим, что $m|ab$ и пусть $(m, a) = d$, $m = dm_1$ и $a = da_1$. Отсюда $(dm_1, da_1) = d$, т. е. $(m_1, a_1) = 1$ и $(m_1b, a_1b) = b$. Очевидно и соотношения $m_1|m_1b$ и $m_1|a_1b$ следуют из условия $m|ab$ после деления на d. В таком случае m_1 должно быть делителем и числа b. С помощью обозначения $b = r - s$ результат может быть представлен в виде:

> Если $ar \equiv as \pmod{m}$, $(a, m) = d$ и $m = dm_1$, $a = da_1$, то имеет место конгруэнтность $r \equiv s \pmod{m}$.

Если $(a, m) = 1$, то получится конгруэнтность $r \equiv s \pmod{m}$.

Результат формулируется следующим образом:

Любая конгруэнтность может быть сокращена на относительное неделимое число, являющееся неделимым относительно модуля. Если конгруэнтность сократить на число, не являющееся неделимым относительно модуля, то правильная конгруэнтность получится лишь в том случае, если поделить модуль на НБОД данного числа и модуля.

Сокращенная система остатков. Как было упомянуто выше, НБОД любого из конгруэнтных между собой чисел по модулю m и числа m — одно и то же число. Поэтому, если некоторое число является неделимым относительно числа m, то все числа, конгруэнтные данному числу, также будут неделимыми относительно числа m. Такие числа считаются важными специальными числами. Если из некоторой полной системы остатков выбрать числа, являющиеся неделимыми относительно модуля, то получится так называемая *сокращенная система остатков.* Итак, количество элементов не зависит от самой сокращенной системы остатков. Рассмотрим полную систему остатков, состоящую из чисел от 1 до m. Пусть из этой системы составляется сокращенная система остатков. На основе определения, из указанных чисел следует выбирать те, которые являются неделимыми относительно m. Количество таких чисел составляет $\varphi(m)$, где φ обозначает функцию Эйлера, т. е. любая сокращенная система остатков имеет $\varphi(m)$ элементов. Далее, эти числа, будучи различными элементами некоторой системы остатков, попарно неконгруэнтны. Наконец, по определению сокращенной системы остатков эти числа являются относительными неделимыми числами по отношению к модулю m. Легко убедиться в том, что этими тремя свойствами характеризуются только сокращенные системы остатков, т. е.:

Некоторые числа образуют сокращенную систему остатков по модулю m тогда и только тогда, если их число составляет $\varphi(m)$; они являются попарно неконгруэнтными по модулю m и каждое из них неделимо относительно m.

Мы уже видели, что некоторая сокращенная система остатков имеет такие свойства. Предположим, что в отношении некоторых чисел настоящие три условия выполняются. Поскольку эти числа являются попарно неконгруэнтными, при делении их на m в результатах оказываются различные остатки. Если к этим числам прибавить все остальные остатки, получающиеся при делении на m, то, видимо, в таком случае получится полная система остатков.

Пусть из полученной таким образом полной системы остатков выбирается некоторая сокращенная система остатков, т. е. выбираются все числа, являющиеся неделимыми относительно m. Учитывая, что и исходные числа были такими, они присутствуют среди элементов полученной таким образом сокращенной системы остатков. Количество элементов в полученной сокращенной системе остатков составляет $\varphi(m)$, т. е. сколько было задано первоначально. Следовательно, в сокращенной системе остатков не могут присутствовать другие числа, кроме предварительно заданных и, таким образом, они образуют действительно сокращенную систему остатков. Пользуясь настоящим результатом, нетрудно получить новую сокращенную систему остатков из заданной:

Если $r_1, r_2, \ldots, r_{\varphi(m)}$ представляют собой сокращенную систему остатков по модулю m и $(a, m) = 1$, то $ar_1, ar_2, \ldots, ar_{\varphi(m)}$ также представляет собой сокращенную систему остатков по модулю m.

ТЕОРЕМА ЭЙЛЕРА

Пусть рассматривается полученная в предыдущей части сокращенная система остатков $ar_1, ar_2, \ldots, ar_{\varphi(m)}$. Каждый элемент в этой системе является неделимым относительно m и поэтому конгруэнтным по отношению к некоторому элементу исходной сокращенной системы остатков $r_1, r_2, \ldots, r_{\varphi(m)}$. Учитывая, что вышеприведенные произведения являются попарно неконгруэнтными, каждое из них является конгруэнтным такому элементу исходной сокращенной системы остатков, который не совпадает с другим конгруэнтным элементом. Пусть

$$\begin{aligned} ar_1 &\equiv r_{i_1} \pmod{m} \\ ar_2 &\equiv r_{i_2} \pmod{m} \\ &\vdots \\ ar_{\varphi(m)} &\equiv r_{i_{\varphi(m)}} \pmod{m}. \end{aligned}$$

В правой стороне этих конгруэнтностей приведены все элементы исходной сокращенной системы остатков, только не в первоначальном порядке. Если перемножить данные конгруэнтности, то опять получится правильная конгруэнтность. Запишем ее, расставляя при этом числа, стоящие в правой стороне, в исходный порядок:

$$a^{\varphi(m)} r_1 r_2 \ldots r_{\varphi(m)} \equiv r_1 r_2 \ldots r_{\varphi(m)} \pmod{m}.$$

Каждый сомножитель произведения $R = r_1 r_2 \ldots r_{\varphi(m)}$ является неделимым относительно m, т. е. имеет место и соотношение $(R, m) = 1$. Но в таком случае вышеприведенная конгруэнтность может быть поделена на R, из чего вытекает соотношение

$$a^{\varphi(m)} \equiv 1 \pmod{m},$$

и получается теорема Эйлера:

Если $(a, m) = 1$, то действительно соотношение $a^{\varphi(m)} \equiv 1 \pmod{m}$.

ТЕОРЕМА ФЕРМА

Пусть $m = p$ представляет собой неделимое число. Тогда $\varphi(p) = p - 1$. Этот частный случай является *теоремой Ферма: Если* $p \nmid a$, *то* $a^{p-1} \equiv 1 \pmod{p}$ *(p — неделимое число).*

Здесь условие $(p, a) = 1$ заменяется эквивалентным ему условием $p \nmid a$. Если умножить последнюю конгруэнтность на a, то получится конгруэнтность $a^p \equiv a \pmod{p}$, которая имеет место и если $p \mid a$, так как в этом случае как a^p, так и a являются конгруэнтными 0 по модулю p. Таким образом, получается конгруэнтность $a^p \equiv a \pmod{p}$ из теоремы Ферма. Но данное утверждение правильно и в том случае, если известно, что $a^p \equiv a \pmod{p}$ и $p \nmid a$, то можно сократить на a и получится конгруэнтность $a^{p-1} \equiv 1 \pmod{p}$. Последней форме чаще отдают предпочтение, так как не следует ставить никакие условия для a:

Теорема Ферма: Если p представляет собой неделимое число, то имеет место конгруэнтность $a^p \equiv a \pmod{p}$.

Данная теорема известна также как «маленькая теорема Ферма» (ниже будет представлена «большая теорема Ферма»).

Как применение теоремы Эйлера будет показано, что каждому числу, неделимому относительно 10, может быть найдено соответствующее «правило делимости». (Конечно, правило не всегда настолько просто, как при числах 9 и 11.) Если $(10, m) = 1$, то по теореме Эйлера имеет место конгруэнтность $10^{\varphi(m)} \equiv 1 \pmod{m}$, из чего можно убедиться в том, что если разделить число, начиная с последнего знака, на интервалы длиной в $\varphi(m)$ знаков, то сумма этих $\varphi(m)$-значных чисел при делении на m дает такой же остаток, как и исходное число. Например, в случае $m = 7$ число может быть разделено на $\varphi(7) = 6$-значные группы. (В предыдущей главе числа были разделены на 3-значные группы, но они брались с чередованием знаков $+$ и $-$.) В некоторых случаях можно добиться результата и при делении числа на более короткие интервалы, чем $\varphi(m)$. Например, в случае $m = 37$ не нужно брать числа длиной в 36 знаков, так как $10^3 = 1000 = 999 + 1 = 27 \cdot 37 + 1$, и из этого следует правильность конгруэнтности $10^3 \equiv 1 \pmod{37}$. Таким образом, достаточно брать трехзначные числа.

(ЛИНЕЙНЫЕ) КОНГРУЭНТНОСТИ ПЕРВОГО ПОРЯДКА

Теорема Эйлера обеспечивает принципиальную возможность проводить деление. В алгебре деление осуществяется с помощью уравнения первого порядка. Так как операция деления числа b на a равноценна поиску такого числа x, в отношении которого имеет место равенство $ax = b$, то следует найти решение уравнения первого порядка типа $ax = b$. Или:

Конгруэнтность вида $ax \equiv b \pmod{m}$ называется линейной конгруэнтностью (первого порядка) при заданных значениях a и b .

Решение линейных конгруэнтностей. При решении линейных конгруэнтностей различают два случая.

1. $(a, m) = 1$.

Если в отношении некоторого значения x имеет место конгруэнтность $ax \equiv b \pmod{m}$, то после умножения на выражение $a^{\varphi(m)-1}$ получится конгруэнтность $a^{\varphi(m)}x \equiv ba^{\varphi(m)-1} \pmod{m}$. На основе теоремы Эйлера это означает, что может выполниться только соотношение $x \equiv ba^{\varphi(m)1} \pmod{m}$.

С другой стороны, если данное условие выполняется при значении x, то это действительно будет решением заданной конгруэнтности, так как в таком случае в соответствии с утверждением выполняется соотношение $ax \equiv aba^{\varphi(m)-1} \equiv ba^{\varphi(m)} \equiv b \pmod{m}$. Итак, в области конгруэнтностей получится не единственное решение, как в области уравнений первого порядка, а бесконечное количество решений, т. е. можно найти бесконечное количество чисел, являющихся конгруэнтными некоторому числу. Но в обдасти конгруэнтностей числа, являющиеся конгруэнтными друг другу, считаются тождественными между собой:

если $(a, m) = 1$, то единственным решением конгруэнтности $ax \equiv b \pmod{m}$ является $x \equiv ba^{\varphi(m)-1} \pmod{m}$.

2. $(a, m) = d \neq 1$.

В этом случае конгруэнтность $ax \equiv b \pmod{m}$ обычно не имеет решения. Если значение c является решением данной конгруэнтности, то на основе соотношений $d \mid ac$ и $d \mid m$ получается соотношение $d \mid (ac, m)$. Но НБОД конгруэнтных чисел и их модули совпадают. Таким образом, $d \mid (b, m)$, а из этого вытекает правильность соотношения $d \mid b$. В конечном итоге:

Необходимым условием разрешимости конгруэнтности $ax \equiv b \pmod{m}$ является $(a, m) \mid b$.

Покажем, что настоящее условие для разрешимости вышеприведенной конгруэнтности является не только необходимым, но и *достаточным*. Иными словами, при выполнении данного условия конгруэнтность, приведенную выше, можно решить.

Запишем значения a, d и m в виде $a = da_1$, $b = db_1$ и $m = dm_1$. Тогда при делении исходной конгруэнтности на d (делится и модуль!) получится конгруэнтность $a_1x \equiv b_1 \pmod{m_1}$. Если x является решением исходной конгруэнтности, то оно будет и решением последней конгруэнтности. Но конгруэнтность $a_1x \equiv b_1 \pmod{m_1}$ на основе соотношения $(a_1, m_1) = 1$, как это было представлено в 1-ом случае, является разрешимой. Если r представляет собой одно из решений, то имеет место конгруэнтность $a_1r \equiv b_1 \pmod{m_1}$, или в иной форме: $m_1 \mid (a_1r - b_1)$, т. е. $dm_1 \mid d(a_1r - b_1)$.

Учитывая соотношения $dm_1 = m$ и $a(a_1r - b_1) = da_1r - db_1 = ar - b$, получится, что $m \mid (ar - b)$, т. е. действительно $ar \equiv b \pmod{m}$. Таким образом, достаточность условия доказана. Осталась неразрешенной проблема определения количества решений исследуемой конгруэнтности. Различными считаются лишь неконгруэнтные решения. Конгруэнтность $a_1x \equiv b_1 \pmod{m_1}$ имеет единственное решение по модулю m_1. Но неконгруэнтные решения не все являются неконгруэнтными по модулю m. Если r представляет собой решение по модулю m_1, то все решения имеют вид: $r, r + m_1, r + 2m_1, \ldots$ и, кроме того, выражения $r - m_1, r - 2m_1, \ldots$ также представляют собой решения. Вопрос в том, когда два таких решения являются неконгруэнтными по модулю m. Поскольку $r + dm_1 = r + m$ является конгруэнтным числу r по модулю m, то каждое решение исследуемой конгруэнтности по модулю m является конгруэнтным некоторому из чисел

$$r,\ r + m_1, \ldots, r + (d - 1)m_1.$$

Но эти числа уже являются неконгруэнтными по модулю m. Ведь пусть двое из них выбираются; путем вычитания меньшего из большего числа получится некоторое кратное числа m_1 (в крайнем случае $(d - 1)$-кратное). Т. е. исследуемая разность является положительным числом, меньшим m, никогда не являющимся делимым на m из-за положительности числа m. В конечном итоге получается такой результат:

> Пусть будет действительно соотношение $(a, m) = d$. Конгруэнтность $ax \equiv b \pmod{m}$ является разрешимой тогда и только тогда, если действительно соотношение $d \mid b$. Если данное условие выполняется и r является решением данной конгруэнтности, то числами $r, r + m_1, \ldots, r + (d - 1)m_1$ определяются все решения данной конгруэнтности, где $m = dm_1$.

СИСТЕМЫ ОДНОВРЕМЕННО ДЕЙСТВИТЕЛЬНЫХ КОНГРУЭНТНОСТЕЙ

Если для определения одной переменной задано уравнение первого порядка, то нет необходимости задавать больше уравнений, так как, если решением нового уравнения является решение исходного, тогда новое уравнение является лишним, а если решение нового уравнения не является решением исходного, то в таком случае получится противоречие. При конгруэнтностях положение другое. Если записывать конгруэнтности с различными модулями для определения одной переменной, то этими конгруэнтностями определяются различные свойства этой переменной и возможно, что конгруэнтности имеют одно общее решение. Например, если о некотором числе известно, что оно является нечетным и при его делении на 5 оказывается остаток 2, то можно сделать вывод, что если делить данное число на 10, то получится остаток 7, т. е. данное число заканчивается цифрой 7.

Если для определения некоторой переменной задаются несколько конгруэнтностей с различными модулями, то такая система называется системой одновременно действительных конгруэнтностей.

Решение систем одновременно действительных конгруэнтностей. Некоторая система одновременно действительных конгруэнтностей очевидно *может иметь* решение только в том случае, если каждая конгруэнтность в отдельности является разрешимой. Но это еще не означает разрешимости системы одновременно действительных конгруэнтностей. Но если отдельные конгруэнтности данной системы являются разрешимыми, то их можно решать отдельно и, следовательно, каждая система однвременно действительных конгруэнтностей может быть задана в виде:

$$x \equiv a_1 \pmod{m_1}, \ldots, x \equiv a_k \pmod{m_k}.$$

Возможность неразрешимости такой системы одновременно действительных конгруэнтностей иллюстрируется следующим примером: пусть задаются конгруэнтности $x \equiv 1 \pmod 2$ и $x \equiv 2 \pmod 4$. Данная система одновременно действительных конгруэнтностей не имеет решения, так как первая конгруэнтность удовлетворяется только при нечетных числах, а вторая — при четных числах. Оказывается достаточным исследовать проблемы систем одновременно действительных конгруэнтностей, состоящих лишь из двух конгруэнтностей, так как тогда возможно решать системы, состоящие из любого количества конгруэнтностей, если решение осуществляется последовательно, (при условии, что система является разрешимой). Итак, пусть рассматривается система одновременно действительных конгруэнтностей вида

$$x \equiv a \pmod m, \qquad x \equiv b \pmod k.$$

По первой когруэнтности x имеет вид: $x = my + a$, где y — некоторое целое число. Если подставить данное выражение во вторую конгруэнтность и применить простые преобразования, то получится конгруэнтность вида

$$my \equiv b - a \pmod{k},$$

которая является линейной конгруэнтностью в отношении y. Необходимым и достаточным условием разрешимости данной конгруэнтности является выполнение условия $(m, k) \mid b - a)$. Данное условие в то же время представляет собой необходимое и достаточное условие разрешимости вышеприведенной системы одновременно действительных конгруэнтностей. Остается только один вопрос: какой вид имеет решение конгруэнтности. Пусть значения $d, m, k, b - a$ записываются в виде $d = (m, k)$, $m = dm_1$, $k = dk_1$ и $b - a = dc$. Подставляя данные выражения в полученную конгруэнтность и сокращая на d, получится конгруэнтность вида

$$m_1 y \equiv c \pmod{k_1},$$

которая является разрешимой и ее решением является выражение $y_0 \equiv cm_1^{\varphi(k_1)-1} \pmod{k_1}$. Итак, все числа, представляющие собой решение данной конгруэнтности, могут быть записаны в виде $y_0 + tk_1 = y$, где t — любое целое число. Если полученное решение подставить обратно в выражение $x = my + a$, получится выражение $x = my_0 + mtk_1 + a = (my_0 + a) + tmk_1$. Поскольку здесь t может быть любым целым числом, то решение получено относительно модуля mk_1. Но $mk_1 = [m, k]$; поэтому решением системы одновременно действительных конгруэнтностей является выражение:

$$\boxed{x \equiv mcm_1^{\varphi(k_1)-1} + a \pmod{[m, k]}.}$$

КОНГРУЭНТНОСТИ ВЫСШИХ ПОРЯДКОВ С НЕДЕЛИМЫМ МОДУЛЕМ

Аналогично понятию уравнений более высоких порядков, можно ввести понятие конгруэнтностей более высоких порядков.

Пусть рассматривается многочлен вида $f(x) = a_n x^n + \ldots + a_1 x + a_0$ с целыми коэффициентами.

Если вюполняется условие $a \not\equiv 0 \pmod{m}$, то конруэнтность $f(x) \equiv 0 \pmod{m}$ называется конгруэнтностью n-го порядка.

По определению целое число a является корнем (решением) многочлена $f(x)$ по модулю m, если имеет место конгруэнтность $f(a) \equiv 0 \pmod{m}$.

Два корня некоторой конгруэнтности называются различными в том случае, когда они являются неконгруэнтными по модулю m. В следующей части рассматричаются только конгруэнтности с неделимым модулем. Сначала будет показано, что при модуле p каждая конгруэнтность может быть приведена к конгруентности, имеющей по крайней мере порядок $(p-1)$. Для выражений $f(x)$ и $x^p - x$ с целыми коэффициентами может быть осуществлена операция деления с остатком (см. главу «Многочлены»), т. е. могут быть найдены такие многочлены $q(x)$ и $r(x)$, в отношении которых выражение $f(x)$ может быть записано в виде $f(x) = q(x)(x^p - x) + r(x)$ и $r(x)$, или многочлен, имеющий по крайней мере порядок $(p - 1)$, или оно тождественно равняется 0. В таком случае для любого числа a действительно выражение:

$$f(a) \equiv q(a)(a^p - a) + r(a) \pmod p.$$

Но на основе теоремы Ферма $a^p - a \equiv 0 \pmod p$, отсюда — правильность выражения $q(a)(a^p - a) \equiv 0 \pmod p$, наконец, получится:

$$f(a) \equiv r(a) \pmod p.$$

Таким образом, корни многочленов $f(x)$ и $r(x)$ совпадают по модулю p и порядок многочлена $r(x)$ — по крайней мере $(p - 1)$, как утверждалось. (Отметим, однако, что в общем случае мультипликативность корней двух многочленов не одинакова. Например, $x = 0$ является p-кратным корнем многочлена x^p, и только однократным корнем соотвествующего ему многочлена x. Понятие мультипликативности приведено в главе «Алгебра».

Количество корней многочлена при неделимом модуле. Ниже будет показано, что при неделимом модуле некоторый многочлен имеет по крайней мере такое количество корней, какой у него порядок.

Пусть проводится операция деления с остатком в отношении *многочлена* $f(x)$ и выражения $(x - a_1)$ (см. главу «Многочлены»), тогда получится соотношение:

$$f(x) = (x - a_1) f_1(x) + f(a_1).$$

Если здесь имеет место конгруэнтность $f(a_1) \equiv 0 \pmod p$, то действительно соотношение

$$f(x) \equiv (x - a_1) f_1(x) \pmod p.$$

Если многочлен $f_1(x)$ имеет корень a_2 по модулю p, то повторяя процесс, приведенный выше, получится:

$$f(x) \equiv (x - a_1)(x - a_2) f_2(x) \pmod p;$$

наконец, получается конгруэнтность:

$$f(x) \equiv (x - a_1)(x - a_2) \ldots (x - a_r) f_r(x) \pmod p,$$

где многочлен $f_r(x)$ уже не имеет корня по модулю p. Поскольку порядок произведения равен сумме порядков сомножителей, порядком *многочлена* $f(x)$ является по крайней мере r. Предположим сейчас, что a является корнем многочлена $f(x)$ по модулю p, т. е. $f(a) \equiv 0 \pmod p$. Из этого следует:

$$(a - a_1)(a - a_2) \ldots (a - a_r) f_r(a) \equiv 0 \pmod p.$$

Это означает, что произведение на левой стороне является делимым на неделимое число p. Из этого следует, что p является делителем и некотогого сомножителя данного произведения. Данным сомножителем не может являться $f_r(a)$ по условию для многочлена $f_r(x)$, итак для некоторого индекса i выполняется условие $p \mid (a - a_i)$, т. е. $a \equiv a_i \pmod p$. На основе проведенного рассуждения любой из корней многочлена $f(x)$ равен по модулю p некоторому из чисел $a_1, a_2, \ldots, a_r$ и, таким образом, многочлен $f(x)$ имеет по крайней мере такое количество корней, какой у него порядок. Данная теорема в общем случае не является правильной. По модулю 6 корнями многочлена $(x-1)(x-2)$ являются 1 и 2. Но кроме них его корнями являются и 4 и 5, так как действительны соотношения $(4-1)$, $(4-2) \equiv$ $\equiv 3 \cdot 2 \equiv 0 \pmod 6$ и $(5-1)(5-2) \equiv 4 \cdot 3 \equiv 0 \pmod 6$.

ТЕОРЕМА УИЛЬСОНА

Пользуясь результатом, представленным в предыдущей части, можно доказать теорему Уильсона:

При любом неделимом числе p имеет место конгруэнтность $(p-1)! \equiv$ $\equiv -1 \pmod p$.

(Через $n!$ обозначается произведение всех целых чисел от 1 до n.)

Пусть рассматривается конгруэнтность вида:

$$(x^{p-1} - 1) - (x-1)(x-2) \ldots [x - (p-1)] \equiv 0 \pmod p.$$

На основе теоремы Ферма корнями настоящей конгруэнтности являются $1, 2, \ldots, p-1$, т. е. она имеет $p-1$ корней. Очевидно, что она является конгруэнтностью по крайней мере $(p-1)$-го порядка, так как оба ее члена являются многочленами именно порядка $(p-1)$. Учитывая, что оба ее члена начинаются выражением x^{p-1}, в разности не будет члена в степени $(p-1)$, т. е. представленная выше конгруэнтность является конгруэнтностью по крайней мере порядка $(p-2)$. Но в таком случае она может иметь $p-1$ корней лишь тогда, если она равна тождественно 0. Учитывая, в частности, постоянный член и перегруппируя отдельные члены, имеем: $-1 \equiv (-1)^{p-1}$ $(p-1)! \pmod p$. Если p — число нечетное, то получится доказуемая конгруэнтность, так как -1 в степени четного числа дает в результате $+1$. Четным неделимым числом является единственное 2, но в данном случае теорема Уильсона очевидно будет действительной.

Отметим, что теорема Уильсона «характерна» для неделимых чисел, т. е. если выполнится условие $(m-1)! \equiv -1 \pmod m$, то число m должно быть неделимым числом. Если $m = 4$, то имеет место соотношение $(m-1)! \equiv 2$ $\pmod m$. Во всех остальных случаях, когда m равняется некоторому сложному числу, большему 4, выполняется условие $(m-1)! \equiv 0 \pmod m$.

Если m представляет собой неделимое число, то число $(m-1)!$ является произведением элементов некоторой сокращенной системы остатков. Соответственно, теорема Уильсона может быть обобщена для любого модуля и, таким образом, рассматривается произведение элементов некоторой сокращенной системы остатков.

Можно доказать, что произведение элементов некоторой сокращенной системы остатков является конгруэнтным по модулю m числу (-1) (элементы берутся по модулю m), если $m = 4$, или если m является степенью некоторого нечетного неделимого числа; в любом другом случае произведение будет конгруэнтным числу 1 по модулю m. Например, при $m = 8$, $1 \cdot 3 \cdot 5 \cdot 7 \equiv 1 \pmod{8}$.

(КВАДРАТИЧНЫЕ) КОНГРУЭНТНОСТИ ВТОРОГО ПОРЯДКА

Довольно много проблем возникает при исследовании конгруэнтностей более высокого порядка при сложном модуле. Поэтому и конгруэнтности второго порядка исследуются здесь только при неделимых модулях. Аналогично тому, как это делается при уравнениях второго порядка, каждая конгруэнтность может быть приведена к конгруэнтности вида $x^2 \equiv a \pmod{p}$. Решение конгруэнтностей такого типа осуществляется фактически с помощью проб. Если, например, следует решить конгруэнтность $x^2 \equiv 6 \pmod{7}$, то вместо x подставляются все числа от 0 до 6. Остатками полученных квадратов являются 0, 1, 4, 2, 2, 4 и 1, соответственно, т. е. приведенная конгруэнтность не может быть решена. Но если бы модуль представлял собой большое число, то пришлось бы испробовать множество чисел, и можно было бы убедиться в том, что исследуемая конгруэнтность не имеет решения. Поэтому математики искали такой метод, с помощью которого относительно просто можно было бы убедиться, существует ли решение данной квадратичной конгруэнтности или нет.

Исследование разрешимости конгруэнтностей второго порядка. Пусть рассматривается конгруэнтность $x^2 \equiv a \pmod{p}$ при заданном значении a. Если выполняется условие $a \equiv 0 \pmod{p}$, то очевидно выполняется и условие $x \equiv 0 \pmod{p}$; в этом случае конгруэнтность, приведенная выше, имеет единственное решение. Пусть этот случай не рассматривается и предположим, что вышеприведенная конгруэнтность является разрешимой; пусть имеет некоторое решение $c \cdot c^2 \equiv a \pmod{p}$. Но в таком случае на основе соотношения $(p-c)^2 = p^2 - 2pc + c^2$ имеет место и когруэнтность $(p-c)^2 \equiv c^2 \pmod{p}$ также и при нечетном значении p (итак, если $p \neq 2$): $c \not\equiv p - c \pmod{p}$. В этом случае вышеприведенная конгруэнтность имеет два решения (как мы видели, количество решений не может быть больше чем 2.) Следовательно, при конгруэнтности $x^2 \equiv a \pmod{p}$ различаются три случая. Если $a \equiv 0 \pmod{p}$, то конгруэнтность имеет одно решение; в других случаях конгруэнтность или не имеет решения, или имеет два неконгруэнтных решения по модулю p.

Квадратичный остаток и квадратичный неостаток. Число a, неделимое на p, называется *квадратичным остатком по модулю p*, если конгруэнтность $x^2 \equiv a \pmod{p}$ является разрешимой, а во всех случаях оно называется *квадратичным неостатком*.

В дальнейшем случай $p = 2$ (как неинтересный) не рассматривается.

Поскольку любой квадратичный остаток является конгруэнтным квадрату двух неконгруэнтных чисел, то количество квадратичных остатков составляет половину количества элементов некоторой сокращенной системы остатков, т. е. $\frac{p-1}{2}$. Учитывая, что как квадратичные остатки, так и квадратичные неостатки неделимы относительно модуля, общее их количество составляет всего $p-1$ и, таким образом, количество квадратичных неостатков также $\frac{p-1}{2}$. Иными словами:

половина элементов некоторой сокращенной системы остатков представляет собой кваратичные остатки, а другая половина — квадратичные неостатки. Свойство некоторого числа, определяющее, является ли данное число квадратичным остатком или нет, называется квадратичным знаком данного числа.

Квадратичный знак произведения. Произведение квадратичных остатков также является квадратичным остатком.

Предположим, что оба числа a и b являются квадратичными остатками по модулю p, т. е. существуют такие неделимые относительно p числа c и d, в отношении которых выполняются условия $c^2 \equiv a \pmod p$ и $d^2 \equiv b \pmod p$. В таком случае для числа cd, также неделимого относительно p, в соответствии с утверждением, имеет место соотношение $(cd)^2 \equiv c^2d^2 \equiv ab \pmod p$.

Произведение некоторого квадратичного остатка и некоторого квадратичного неостатка является квадратичным неостатком.

Предположим, что a является квадратичным остатком по модулю p и пусть умножаются элементы некоторой сокращенной системы остатков на a. Произведения опять представляют собой сокращенную систему остатков. По предыдущим результатам, если некоторый квадратичный остаток умножается на a, опять получается квадратичный остаток. Т.е. умножая $\frac{p-1}{2}$ из исходной сокращенной системы остатков на a, получим $\frac{p-1}{2}$ квадратичных остатков полученной сокращенной системы остатков, т. е. получены все квадатичные остатки. На основе такого рассуждения, если умножить некоторый квадратичный неостаток на a, то нельзя получить квадратичный остаток, и поскольку ни одно из этих произведений не может быть делимым на p, результатами данных произведений являются квадратичные неостатки (в соответствии с утверждением, представленным выше).

Произведение двух квадратичных неостатков являетя квадратичным остатком.

Посмотрим квадратичный неостаток a и пусть умножаются все элементы некоторой сокращенной системы остатков на a. Умножая квадратичный остаток, получаем квадратичный неостаток. Таким образом, перемножив $\frac{p-1}{2}$ квадратичных остатков, получаются все квадратичные неостатки, поэтому, если умножить квадратичный неостаток на a, то может быть получен уже только квадратичный остаток.

ОПРЕДЕЛЕНИЕ КВАДРАТИЧНОГО ЗНАКА

В отношении операции произведения квадратичный остаток и квадратичный неостаток характеризуются подобно числам $+1$ и -1:

остаток · остаток = остаток	$(+1)\cdot(+1) = +1$
остаток · неостаток = неостаток	$(+1)\cdot(-1) = -1$
неостаток · неостаток = остаток	$(-1)\cdot(-1) = +1$

Для облегчения вычислений к квадратичному остатку предписывается число $(+1)$, а к квадратичному неостатку — число (-1). Учитывая, что квадратичный знак некоторого числа зависит и от модуля, при предписании следует указать и модуль. Все это осуществляется с помощью так называемого *символа Лежандра*:

$$\left(\frac{a}{p}\right) = +1, \quad \text{если } a \text{ является квадратичным остатком по модулю } p,$$

$$\left(\frac{a}{p}\right) = -1, \quad \text{если } a \text{ является квадратичным неостатком по модулю } p.$$

[Знак $\left(\frac{a}{p}\right)$ читается как «a на p; символ Лежандра», или «a деленное на p».]

Итак, на основе приведенных выше сведений, имеет место соотношение $\left(\frac{a\cdot b}{p}\right) = \left(\frac{a}{p}\right)\left(\frac{b}{p}\right)$. Данное соотношение остается действительным и в том случае, если при $p \mid a$ символ Лежандра определяется в виде $\left(\frac{a}{p}\right) = 0$.

Эти три определения могут быть соединены в одно общее понятие:

пусть m представляет собой количество решений конгруэнтности $x^2 \equiv a \pmod{p}$.

Тогда имеет место соотношение: $\left(\frac{a}{p}\right) = m - 1$.

Доказанная выше мультипликативность символа Лежандра и основная теорема чисел указывают на то, что достаточно определить символ $\left(\frac{q}{p}\right)$ в том случае, если q представляет собой неделимое число. Но исследование целесообразно проводить в более общем кругу. Если, например, следует выяснить, является ли число 97 квадратичным остатком по модулю 101, то целесообразно вместо 97 исследовать конгруэнтное ему число -4. Таким образом, и в общем случае целесообразно допускать присутствие отрицательных чисел, поэтому следует определить и понятие символа Лежандра $\left(\frac{-1}{p}\right)$.

Доказательство результатов в области символа Лежандра довольно продолжительно и сложно, поэтому здесь приводятся лишь соответствующие теоремы.

Если p — нечетное неделимое число, то имеет место соотношение $\left(\frac{-1}{p}\right) = (-1)^{\frac{p-1}{2}}$. Число -1 является квадратичным остатком по модулю p, если модуль имеет вид $p = 4k + 1$ и квадратичным неостатком, если p имеет вид $p = 4k - 1$.

Если p представляет собой нечетное неделимое число, то действительно соотношение: $\left(\frac{2}{p}\right) = (-1)^{\frac{p^2-1}{8}}$. Число 2 является квадратичным остатком по модулю p, если p имеет вид или $p = 8k + 1$, или $p = 8k - 1$; и квадратичным неостатком, если p имеет вид или $p = 8k + 3$ или $p = 8k - 3$.

Теорема обратимости относится к нечетным неделимым числам. Сама теорема уже была представлена Эйлером, но он не сумел доказать ее. С тех пор найдено большое число доказательств. Независимо от Эйлера теорема была представлена и Лежандром; он доказал часть теоремы, хотя и очень сложным путем. Первое доказательство теоремы принадлежало Гауссу, который представил шесть различных доказательств теоремы. Самым известным считается третье по временному порядку доказательство Гаусса, в котором применяется так называемая лемма Эйлера и Гаусса.

Если p и q — различные нечетные неделимые числа, то выполняются условия:

$$\left(\frac{q}{p}\right)\left(\frac{p}{q}\right) = (-1)^{\frac{p-1}{2}\cdot\frac{q-1}{2}}, \text{ т. е. } \left(\frac{q}{p}\right) = \left(\frac{p}{q}\right)(-1)^{\left(\frac{p-1}{2}\right)\left(\frac{q-1}{2}\right)}.$$

Теорему можно представить и в другом виде:

$\left(\frac{q}{p}\right) = \left(\frac{p}{q}\right)$, если хотя бы одно из чисел p и q имеет вид: $4k + 1$ и

$\left(\frac{q}{p}\right) = -\left(\frac{p}{q}\right)$, если оба числа p и q имеют вид: $4k - 1$.

Пусть определяется разрешимость конгруэнтности $x^2 \equiv 127 \pmod{157}$. Сначала следует определить, является ли модуль неделимым числом. Поскольку 157 меньше, чем 13 в квадрате, то делимость проверяется только при 2, 3, 5, 7 и 11, а так как число 157 является неделимым относительно этих чисел, то 157 представляет собой неделимое число. Рассмотрим символ Лежандра:

$$\left(\frac{127}{157}\right) = \left(\frac{-30}{157}\right) = \left(\frac{-1}{157}\right)\left(\frac{2}{157}\right)\left(\frac{3}{157}\right)\left(\frac{5}{157}\right).$$

Поскольку присутствует и символ $\left(\frac{2}{p}\right)$, рассмотрим остаток при делении на 8: $157 = 8 \cdot 19 + 5$:

$$\left(\frac{-1}{157}\right) = 1, \qquad \left(\frac{2}{157}\right) = -1, \qquad \left(\frac{3}{157}\right) = \left(\frac{157}{3}\right) \quad \text{и}$$

$$\left(\frac{5}{157}\right) = \left(\frac{157}{5}\right).$$

При делении 157 на 3 остаток будет 1, а на 5 — 2; поэтому $\left(\frac{157}{3}\right) = \left(\frac{1}{3}\right) = 1$ (так как число 1 является всегда квадратичным остатком); $\left(\frac{157}{5}\right) = \left(\frac{2}{5}\right) = -1$, так как число 5 записывается в виде $8k - 3$. В конечном итоге символ Лежандра имеет вид;

$$\left(\frac{127}{157}\right) = 1 \cdot (-1) \cdot 1 \cdot (-1) = 1;$$

следовательно, конгруэнтность, приведенная выше, разрешима. (Действительно, $21^2 \equiv 127 \pmod{157}$).

ПРОСТОЙ КОРЕНЬ

По теореме Ферма, если $p \nmid a$, то выполняется условие $a^{p-1} \equiv \pmod{p}$. Но если a является квадратичным остатком по модулю p, т. е. имеется такое значение c, в отношении которого выполняется условие $c^2 \equiv a \pmod{p}$ и при этом $p \nmid c$), то соотношение $a^{\frac{p-1}{2}} \equiv (c^2)^{\frac{p-1}{2}} \equiv c^{p-1} \equiv 1 \pmod{p}$ является правильным, т. е. и $\frac{p-1}{2}$-ая степень числа a является конгруэнтной числу 1 по модулю p.

Если k — наименьшее натуральное число, в отношении которого имеет место соотношение $a^k \equiv 1 \pmod{p}$, то число k называется порядком числа a.

Покажем, что *если k является порядком числа a, то имеет место соотношение $k \mid (p-1)$.*

Если проводить операцию деления с остатком для чисел $p-1$ и k, получится соотношение $p-1=ku+r$, где r — число неотрицательное и меньше k, из чего следует правильность соотношения:

$$1 \equiv a^{p-1} \equiv a^{ku+r} \equiv (a^k)^u(a^r) \equiv 1^u a^r \equiv a^r \pmod p.$$

По определению числа k данное соотношение может иметь место только в том случае, когда выполняется условие $r=0$, т. е. $p-1=ku$, которое равноценно соотношению $k|(p-1)$. Таким же образом можно убедиться и в том, что если k — порядок числа a и выполняется условие $a^s \equiv 1 \pmod p$, то имеет место и соотношение $k \mid s$.

Покажем, что *если выполняется условие $k \mid (p-1)$, то имеется $\varphi(k)$ таких неконгруэнтных по модулю p чисел, порядком которых является k.*

Если выполняется условие $k|(p-1)$, то многочлен $x^{p-1}-1$ может быть разложен в виде: $x^{p-1}-1=(x^k-1)(x^{p-1-k}+x^{p-1-2k}+\ldots+x^{2k}+x^k+1)$. Первый сомножитель многочлена по модулю p имеет по крайней мере k (неконгрэнтных) корней. Но число корней первого сомножителя не может быть меньшим, так как в таком случае число корней произведения могло бы быть по крайней мере равным числу корней двух сомножителей, взятых вместе, и поскольку и второй сомножитель может иметь число корней, равное его порядку, то число корней произведения составило бы меньше, чем $k+(p-1-k)=p-1$. Но с другой стороны, по теореме Ферма произведение имеет именно $p-1$ корней. Следовательно, при $k|(p-1)$ число корней многочлена x^k-1 составляет именно k. Пусть надо вычислить, сколько существует чисел, неконгруэнтных по модулю p, порядком которых является именно k.

Поскольку многочлен x^{k-1} имеет именно k корней, существуют k таких чисел, порядок которых является делителем числа k.

Если следует определить, сколько существует чисел, порядком которых является k, то из этих чисел необходимо удалить те числа, порядок которых меньше, чем k.

Пример: Процесс представляется на примере; в общем случае таким же путем можно достичь результата.

Пусть задается число $k=1200=2^4\cdot 3\cdot 5^2$. Если выполняется условие $a^{1200}=1 \pmod p$ и некоторая из степеней числа a, степень которой меньше чем 1200, является конгруэнтной числу 1 по модулю p, то данная степень является произведением некоторых из неделимых чисел 2, 2, 2, 2, 3, 5, 5 (состоящих по крайней мере из 6 сомножителей). Пусть степень дополняется таким образом, чтобы она состояла из 6 сомножителей, что представляет собой следующее возведение в степень, и поскольку исследуемая степень числа a была конгруэнтна уже числу 1 по модлую p, то полученная таким образом степень конгруэнтна числу 1 по модулю p. Если в произведении имеются 6 сомножителей, то пропущено некоторое (одно) из чисел 2, 3 или 5. Поэтому какое-то из чисел a^{600}, a^{400} и a^{240} является конгруэнтным числу 1 по модулю p $\left(600=\dfrac{1200}{2},\ 400=\dfrac{1200}{3} \text{ и } 240=\dfrac{1200}{5}\right)$. Количество чисел a, имеющих такой характер, составляет 600, 400 и 240, соответственно. Числа, в отношении которых имеют место две из вышеприведенных конгруэнтностей, учитывались дважды. Посмотрим, сколько таких чисел существует. Нетрудно убедиться в том, что, например, оба числа a^{600} и a^{400} являются конгруэнтными числу 1 тогда, если и число a^{200} является конгруэнтным числу 1 (200 является НБОД чисел 600 и 400). Итак, опять следует учитывать числа, в отношении которых одно из чисел a^{200}, a^{120} и a^{80} является конгруэнтным числу 1 по модулю p. [Здесь $120 = (600, 240)$ и $80 = (400, 240)$]. Количество таких чисел составляет 200, 120 и 80, соответственно. Все числа, в отношении которых все три приведенные выше конгруэнтности выполняются, трижды были пропущены, но и трижды учитывались заново. Значит, их опять надо пропустить. Это будут именно те числа (600, 400, 240), 40-ая степень которых конгруэнтна числу 1 по модулю p. Количество таких чисел составляет 40. В конечном итоге количество чисел, являющихся неконгруэнтными по модулю p, имеющих порядок 1200, составляет именно:

$$1200 - 600 - 400 - 240 + 200 + 120 + 80 - 40.$$

Если вынести за скобки число 1200 и образовать произведение, то получится выражение:

$$1200 \cdot \left(1 - \frac{1}{2} - \frac{1}{3} - \frac{1}{5} + \frac{1}{2 \cdot 3} + \frac{1}{2 \cdot 5} + \frac{1}{3 \cdot 5} - \frac{1}{2 \cdot 3 \cdot 5}\right) =$$

$= 1200 \cdot \left(1 - \frac{1}{2}\right) \cdot \left(1 - \frac{1}{3}\right) \cdot \left(1 - \frac{1}{5}\right)$, которое совпадает именно со значением $\varphi(1200)$.

Если число g имеет порядок $p - 1$ по модулю p, то число g называется простым корнем по модулю p.

В соответствии с доказанной выше теоремой имеет место утверждение: *Существует простой корень по модулю p; количество таких корней составляет $\varphi(p - 1)$.*

Если g является простым корнем по модулю p, то числа $g, g^2, \ldots, g^{p-1}$ являются неконгруэнтными, иначе было бы $g^i \equiv g^j \pmod p$, из которого при $i > j$ в отличие от исходного предположения получилось бы соотношение $g^{i-j} \equiv 1 \pmod p$. Таким свойством обладают только простые корни. Если все числа $g, g^2, \ldots, g^{p-1}$ являются неконгруэнтными, то кроме последнего числа ни одно из них не может быть конгруэнтным числу 1 по модулю p, т. е. порядок числа g — именно $p - 1$. Простые корни имеют широкое применение. Например, с их помощью легко могут быть решены конгруэнтности типа $x^r \equiv a \pmod p$. Пусть, например, g является простым корнем по модулю p и одновременно выполняется условие $a \equiv g^s \pmod p$ (случай $p \mid a$ не рассматривается). В этом случае, если искать решение в виде g^y, то получится соотношение $g^{ry-s} \equiv 1 \pmod p$. Данное соотношение является эквивалентным выражению $(p - 1) \mid ry - s$), т. е. решению конгруэнтности первого порядка $ry \equiv s[\mathrm{mod}\ (p - 1)]$.

АДДИТИВНАЯ ТЕОРИЯ ЧИСЕЛ

ЛИНЕЙНЫЕ РАЗЛОЖЕНИЯ

Предметом аддитивной теории чисел являются такие проблемы, при решении которых следует исследовать возможности представления некоторых чисел в виде суммы. Очень трудно ответить на такие вопросы, так как в большинстве случаев в области теории чисел рассматриваются проблемы представления чисел в виде произведения, начиная с проблемы основной теоремы теории чисел. Именно этим и объясняется, что по возможности аддитивное представление приводится к представлению в виде произведения.

Обычно в области проблем аддитивной теории чисел возникают два вопроса: какие числа могут быть представлены в желаемом виде и какое количество различных форм представления существует. Самыми простыми являются так называемые линейные разложения. Пусть заданы фиксированные положительные целые числа $a_1, a_2, \ldots, a_k$. Следует найти решения уравнения

$$(1) \qquad a_1x_1 + a_2x_2 + \ldots a_kx_k = n$$

при выполнении следующих условий. Число n представляет собой натуральное число и решения ищут в зависимости от значения n. Переменные $x_1, x_2, \ldots, x_k$ могут принимать только целочисленные значения; как значения этих переменных допускаются или только положительные числа, или допускается и число 0. Обычно число решений задается в зависимости от n. Предположим, что необходимо найти неотрицательные решения уравнения (1). Тогда выражение $y_i = x_i + 1$ примет положительное значение, и в отношении всех значений y_i имеет место уравнение

$$(2) \qquad a_1y_2 + a_2y_2 + \ldots + a_ky_k = n + a_1 + a_2 + \ldots + a_k.$$

Данное утверждение является правильным и тогда, когда значения y_i представляют собой положительные решения уравнения (2), и значениями $x_i = y_i - 1$ определяются неотрицательные решения уравнения (1). Именно поэтому считается достаточным рассмотрение неотрицательных решений.

Может возникнуть и другая проблема. Пусть задаются, например, конкретные значения $k = n = 3$ и $a_1 = a_2 = a_3 = 1$. В таком случае уравнение (1) примет вид

$$x_1 + x_2 + x_3 = 3.$$

Одно из решений настоящего уравнения: $x_1 = 2$, $x_2 = 1$, $x_3 = 0$; другое его решение: $x_1 = 0$, $x_2 = 1$, $x_3 = 2$. На основе исходного определения решений уравнения настоящие решения являются различными решениями. Но по существу в обоих случаях число 3 было записано в виде $0 + 1 + 2$. Если такие, получаемые путем перестановки одинаковых значений переменных, решения не различаются между собой, то данный процесс поиска корней уравнений называется исследованием *существенно разли-*

чающихся решений уравнения. Исследование существенно различающихся решений намного труднее, чем исследование различных решений. Оба вопроса рассматриваются на примере уравнения типа

$$x_1 + x_2 + \ldots + x_k = n. \tag{3}$$

Сначала рассматриваются различные решения данного уравнения. Здесь целесообразно исследовать положительные решения уравнения. Поэтому вместо уравнения (3) рассматриваются положительные решения уравнения вида

$$y_1 + y_2 + \ldots + y_k = n + k. \tag{4}$$

Вводятся новые переменные $u_1 = y_1$, $u_2 = u_1 + y_2$, $u_3 = u_2 + y_3$, ... $u_{k-1} = u_{k-2} + y_{k-1}$. Поскольку каждое решение y_1 является положительным, то, с одной стороны, выполняется неравенство $0 < u_1 < u_2 < u_3 < \ldots < u_{k-1}$, а с другой стороны — равенство $u_{k-1} = u_{k-2} + y_{k-1} = \ldots = u_2 + y_3 \ldots + y_{k-1} = u_1 + y_2 + y_3 + \ldots + y_{k-1} = y_1 + y_2 + y_3 + \ldots + y_{k-1} = n + k - y_k \leq n + k - 1$. А если числа $u_1, u_2, \ldots, u_{k-1}$ выбираются таким образом, чтобы при этом выполнилось неравенство $0 < u_1 < u_2 < u_3 \ldots < u_{k-1} \leq n + k - 1$, то числа $y_1 = u_1$, $y_2 = u_2 - u_1, \ldots, y_{k-1} = u_{k-1} - u_{k-2}$, $y_k = n + k - u_{k-1}$, с одной стороны, все будут положительны, и с другой, в отношении этих чисел выполняется условие (4). Таким образом, число решений уравнения (4) (и уравнения (3)) зависит от того, какими способами возможно выбрать $k - 1$ различных чисел из чисел 1, 2, ..., $n + k - 1$, что представляет собой комбинаторную задачу, а именно — задачу определения количества комбинаций из $n + k - 1$ элементов по классу $(k - 1)$, определяющегося согласно соотношению $\binom{n+k-1}{k-1}$.

Сама задача и соответствующие соотношения приведены в главе «Комбинаторика». Таким образом, удалось определить решение приведенной выше задачи.

Более трудной является задача определения существенно различающихся решений уравнения (3). Поскольку в этом случае порядок неизвестных параметров не учитывается, можно предполагать, что выполняется неравенство $x_1 \leq x_2 \leq \ldots \leq x_k$. Здесь целесообразно применить новые переменные вида $z_1 = x_1$, $z_2 = x_2 - x_1, \ldots, z_k = x_k - x_{k-1}$. В соответствии с условием, представленным выше, значения переменных z_i неотприцательны и в их отношении выполняется условие

$$z_k + 2z_{k-1} + \ldots + kz_1 = n. \tag{5}$$

В таком случае можно показать, что для значений z_i, удовлетворяющих таким условиям, выражениями вида $x_1 = z_1$, $x_2 = x_1 + z_1, \ldots, x_k = x_{k-1} + z_k$ определяется решение уравнения (3), согласно требованиям, представленным в данном случае. Следовательно, уравнение (3) имеет такое число существенно различающихся решений, которое составляет количество различных решений уравнения (5). Но определение различных решений данного уравнения представляет собой очень сложную задачу.

ГЕНЕРАТОРНАЯ ФУНКЦИЯ

Рассмотрим другое специальное уравнение вида

$$x + 2y + 4z = n. \tag{6}$$

Ставится задача нахождения количества неотрицательных решений настоящего уравнения. Методом, применяемым в данном случае, можно пользоваться и в общем случае для любого уравнения типа (1). Пусть $f(n)$ обозначает количество неотрицательных решений уравнения (6). Вместо того, чтобы определить отдельные значения функции $f(n)$, определим все ее значения одновременно! Используемый здесь прием решения задачи называется *принципом генераторной функции.*

Если $g(n)$ представляет собой любую функцию теории чисел, то степенной ряд

$$G(t) = g(0) + g(1)t + g(2)t^2 + \ldots + g(n)t^n + \ldots$$

называется генераторной функцией функции $g(n)$.

По существу необходимо исследовать проблему сходимости при использовании представленного выше степенного ряда по переменной t, хотя во многих случаях проблему сходимости можно не учитывать. При решении задач, приводимых ниже, генераторная функция всегда сходится в некотором интервале и этим фактом обеспечивается возможность осуществления соответствующих преобразований, необходимых для решения. Применимые понятия представлены в главе «Анализ».

Пусть функция $F(t)$ представляется в виде $F(t) = f(0) + f(1)t + f(2)t^2 + \ldots + f(n)t^n + \ldots$. В соответствии с коэффициентами уравнения (6) рассматриваются следующие три степенных ряда:

$$F_1(t) = 1 + t + t^2 + t^3 + \ldots$$

$$F_2(t) = 1 + t^2 + t^4 + t^6 + \ldots$$

$$F_4(t) = 1 + t^4 + t^8 + t^{12} + \ldots$$

Во всех случаях показатели представляют собой кратные соответствующих коэффициентов. Пусть эти три степенные ряды перемножаются между собой по правилам умножения многочленов по Коши («умножение каждого члена на каждый член»). В произведении получатся члены, каждый из которых состоит из трех сомножителей. Отдельные члены имеют вид: $t^x t^{2y} t^{4z} = t^{x+2y+4z}$, где x, y и z являются неотрицательными целыми числами. Посмотрим, сколько из этих членов равняется выражению t^n. Условие $t^{x+2y+4z} = t^n$ выполняется тогда и только тогда, если выполняется и условие $x + 2y + 4z = n$. Количество таких значений переменных x, y и z зависит от того, сколько различных неотрицательных решений имеет уравнение (6). Т.е. в произведении член t^n наступает именно $f(n)$ раз. Иными словами, в произведении коэффициентом члена t^n является функция $f(n)$. Получился результат:

$$F(t) = F_1(t)\, F_2(t)\, F_4(t)\,.$$

Далее степенные ряды, представленные выше, преобразуются в такой вид, при использовании которого легко определить коэффициенты степенного ряда $F(t)$. На основе формулы суммы бесконечной геометрической прогрессии выражения

$$F_1(t) = \frac{1}{1-t},$$

$$F_2(t) = \frac{1}{1-t^2} \quad \text{и} \quad F_4(t) = \frac{1}{1-t^4}$$

являются правильными. Из этих соотношений для функции $F(t)$ получится выражение:

$$F(t) = \frac{1}{(1-t)^3(1+t)^2(1+t^2)}.$$

Путем применения правил разложения дроби на частные дроби (глава «Анализ»), получится выражение:

$$F(t) = \frac{1}{8}\,\frac{1}{(1-t)^3} + \frac{1}{4}\,\frac{1}{(1-t)^2} + \frac{9}{32}\,\frac{1}{1-t} + \frac{1}{16}\,\frac{1}{(1+t)^2} + \\ + \frac{5}{32}\,\frac{1}{1+t} + \frac{1}{8}\,\frac{1+t}{1+t^2}.$$

Пользуясь правилом исчисления производной дробно-рациональных функций (глава «Анализ»), получим выражение:

$$F(t) = \frac{1}{16}\left(\frac{1}{1-t}\right)'' + \frac{1}{4}\left(\frac{1}{1-t}\right)' + \frac{9}{32}\,\frac{1}{1-t} - \frac{1}{16}\left(\frac{1}{1+t}\right)' + \\ + \frac{5}{32}\,\frac{1}{1+t} + \frac{1}{8}\,\frac{1+t}{1+t^2}.$$

Если провести операцию разложения в степенной ряд для отдельных дробно-рациональных функций, учитывая возможность исчиления производной настоящих степенных рядов отдельно на каждый член, то функция $F(t)$ определяется суммой следующих степенных рядов:

$$\frac{1}{16}[2 + 6t + 12t^2 + \ldots + (n+2)(n+1)t^n + \ldots]$$

$$\frac{1}{4}[1 + 2t + 3t^2 + \ldots + (n+1)t^n + \ldots]$$

$$\frac{9}{32}(1 + t + t^2 + \ldots + t^n + \ldots)$$

$$-\frac{1}{16}[-1+2t-3t^2+\ldots+(-1)^{n-1}(n+1)t^n+\ldots]$$

$$\frac{5}{32}[1-t+t^2-\ldots+(-1)^n t^n+\ldots]$$

$$\frac{1}{8}(1+t)(1-t^2+t^4-\ldots)=\frac{1}{8}(1+t-t^2-t^3+t^4+t^5-\ldots).$$

Из приведенных выше соотношений значение функции $f(n)$ определяется путем суммирования всех коэффициентов членов t^n, присутствующих в отдельных степенных рядах. Единственным затруднением является чередование знаков через каждые два члена в последнем степенном ряду. Коэффициент члена t^n в этом степенном ряду определяется выражением $(-1)^{\frac{n(n-1)}{2}}$. Таким образом, функция определяется согласно:

$$\boxed{f(n)=\frac{(n+2)(n+1)}{16}+\frac{n+1}{4}+\frac{9}{32}+(-1)^n\frac{n+1}{16}+(-1)^n\frac{5}{32}+\frac{1}{8}(-1)^{\frac{n(n-1)}{2}}.}$$

Отметим, что данное выражение — дробь, но оно должо определить целое число.

Пусть $h(n)$ обозначает количество неотрицательных решений уравнения (1). На основе рассмотренного метода, может быть доказано следующее асимптотическое свойство:

$$h(n)\sim\frac{n^{k-1}}{(k-1)!\,a_1a_2\ldots a_k}.$$

С помощью настоящего соотношения асимптотического свойства получатся асимптотические выражения

$$u(n)\sim\frac{n^{k-1}}{(k-1)!}\quad\text{и}$$

$$v(n)\sim\frac{n^{k-1}}{(k-1)!\,k!}$$

для определения количества различных решений $u(n)$ и существенно различающихся решений $v(n)$ уравнения (3).

ПРЕДПОЛОЖЕНИЕ ГОЛЬДБАХА

Оно относится к проблеме линейного характера. По этому предположению каждое четное число может быть записано в виде суммы по крайней мере двух неделимых чисел *(предположение Гольдбаха на четные числа)*. Каждое нечетное число может быть записано в виде суммы по крайней мере трех неделимых чисел *(предположение Гольдбаха на нечетные числа)*. Если предположение Гольдбаха правильно в отношении четных чисел, то правильно и его предположение в отношении нечетных чисел. Если n — нечетное число, то число $n-3$ — четное и если это число записывать в виде $n-3=p+q$, то имеет место соотношение $n=3+p+q$ (p и q — неделимые числа). Из данных предположений, к сожалению, доказано только предположение на нечетные числа (и оно доказано лишь для «достаточно больших» чисел), из чего не следует правильность предположения Гольдбаха в отношении четных чисел.

КВАДРАТИЧНЫЕ РАЗЛОЖЕНИЯ

Рассматривается частный случай квадратичного разложения. Ставится проблема : в каких случаях квадратное число может быть воспроизведено в виде суммы двух квадратных чисел, т. е. проблема определения целых решений x, y, z уравнения вида

$$x^2+y^2=z^2. \tag{1}$$

Учитывая правильность соотношения $(-a)^2=a^2$, можно не рассматривать отрицательные решения данного уравнения. Также можно не рассматривать такие случаи, когда одно из неизвестных переменных принимает значение 0. Следовательно, проблема нахождения положительных решений оказывается достаточной. Если числа x, y, z считаются сторонами некоторого треугольника, то уравнением (1) выражается, что данный треугольник является прямоугольным, в соответствии с теоремой Пифагора. Поэтому решения уравнения (1) называются *пифагорейскими тройными числами*. Данная задача может быть сформулирована и иначе: следует найти такие натуральные числа x, y, z, из которых может быть сконструирован прямоугольный треугольник (с гипотенузой z). Некоторое решение уравнения (1) y x, z называется *основным решением*, если выполняются условия $(x, y)=(y, z)=(z, x)=1$. Покажем, что все решения уравнения (1) могут быть получены путем перемножения элементов основного решения на одно и то же целое число.

Очевидно, применяя метод, предложенный выше, всегда получится решение, так как если числа x, y, z представляют собой любое решение, и t обозначает некоторое целое число, то выполняется условие $(tx)^2+(ty)^2=t^2(x^2+y^2)=t^2z^2=(tz)^2$. Предположим, что решением уравнения (1) является x_1, y_1 и z_1 и применяются обозначения $d=(x_1, y_1, z_1)$ и $x_1=dx$, $y_1=dy$ и $z_1=dz$. В соответствии с условием $(dx)^2+(dy)^2=(dz)^2$ и путем деления на выражение d^2 получится соотношение: $x^2+y^2=z^2$, т. е.

тройные числа x, y, z также представляют собой решение уравнения (1). Покажем, что получено основное решение. Пусть, например, выполняется условие $p \mid (x, y)$. Тогда p является делителем суммы квадратов этих же двух чисел, т. е. числа z^2. Если p — неделимое число, то из этого следует правильность соотношения $p \mid z$, противоречащего условию $(x, y, z) = 1$. Поэтому имеет место соотношение $(x, y) = 1$. Аналогичным образом можно убедиться в правильности соотношений $(y, z) = 1$ и $(z, x) = 1$. В таком случае x и y не могут быть одновременно четными или одновременно нечетными. Ведь если на 4 поделить квадрат некоторого нечетного числа, то в результате деления остаток оказывается 1 и поэтому при делении суммы квадратов двух неделимых чисел на 4 остаток будет 2. Соответственно число z^2 представляет собой четное число, неделимое на 4, а такое условие исключается, так как если некоторое квадратное число является четным, то оно должно быть делимым хотя бы на вторую степень числа 2. В конечном итоге, можно сделать вывод, что одно из чисел x и y является четным, а другое — нечетным и, таким образом, и число z является нечетным. На основе свойства симметричности можно предполагать, что x является четным числом. Пусть уравнение (1) перепишется так:

$$x^2 = z^2 - y^2 = (z + y)(z - y).$$

Поскольку z и y нечетные числа, поэтому как число $z + y$, так и число $z - y$ являются четными, а в отношении целых чисел $\frac{x}{2}$, $\frac{z+y}{2}$ и $\frac{z-y}{2}$ выполняется условие

$$\left(\frac{x}{2}\right)^2 = \left(\frac{z+y}{2}\right)\left(\frac{z-y}{2}\right).$$

Для двух сомножителей, стоящих в правой стороне соотношения, должно выполниться условие относительной неделимости. Если число q было бы их общим делителем, то оно было бы делителем и их суммы, и их разности, т. е. выражений $\frac{z+y}{2} + \frac{z-y}{2} = z$ и $\frac{z+y}{2} - \frac{z-y}{2} = y$, а данное условие может быть выполнено лишь в случае $q = 1$ из-за соотношения $(z, y) = 1$. С помощью основной теоремы теории чисел, произведение двух относительно неделимых чисел может быть квадратным числом только в том случае, когда любое из этих двух чисел представляет собой полный квадрат. Пусть, например, данные числа записываются в виде $\frac{z+y}{2} = u^2$ и $\frac{z-y}{2} = v^2$. Очевидна правильность соотношений $(u, v) = 1$ и $u > v$. Также нетрудно убедиться в том, что одно из чисел u и v является четным, так как если бы оба они были нечетными, то сумма их квадрата представляла бы собой четное число, в то время как число $u^2 + v^2 = z$ является нечетным. Можно предположить, что числа u и v положительны. Итак, путем замены в уравнении, затем извлечения квадратных корней, наконец, умножая полученное соотношение на 2, достигается результат: $x = 2uv$.

Все основные решения уравнения (1) *получаются в виде:*

$$x = 2uv, \qquad y = u^2 - v^2, \qquad z = u^2 + v^2,$$

где u и v — положительные целые числа, в отношении которых выполняются условия $u > v$, $(u, v) = 1$ и одно из этих двух чисел является четным.

Если числа u и v выбраются соответственно вышеприведенным условиям, то имеют место соотношения $x = 2uv$, $y = u^2 - v^2$, $z = u^2 + v^2$, и настоящими выражениями определяются основные решения уравнения (1). Можно показать, что действительно получено решение:

$$z^2 - y^2 = (z + y)(z - y) = (2u^2)(2v^2) = (2uv)^2 = x^2.$$

Нужно показать, что получено именно основное решение. Если выполняется условие $d = (x, y, z)$, то очевидно выполняются и условия $d \mid (x, z + y)$ и $d \mid (x, z - y)$. Но так как имеет место соотношение $(u, v) = 1$, действительны и соотношения $(x, z + y) = (2uv, 2u^2) = 2u(v, u) = 2u$ и $(x, z - y) = (2uv, 2v^2) = 2v(u, v) = 2v$. Из них вытекает правильность соотношения $d \mid (2u, 2v)$, но из правильности соотношения $(2u, 2v) = 2$ получится или $d = 1$ или $d = 2$. Поскольку одно из чисел u и v является четным, число z (и число y также) нечетное, следовательно, число d может быть только нечетным, т. е. $d = 1$. Таким образом получено основное решение. Условием $u > v$ обеспечивается только положительность числа y. Приведем некоторые основные решения:

u	2	3	4	4	5	5	6	6	7	7	7	8	8	8	8
v	1	2	1	3	2	4	1	5	2	4	6	1	3	5	7
x	4	12	8	24	20	40	12	60	28	56	84	16	48	80	112
y	3	5	15	7	21	9	35	11	45	33	13	63	55	39	15
z	5	13	17	25	29	41	37	61	53	65	85	65	73	89	113

Рассмотрим уравнение

(2) $$x^2 + y^2 = n$$

и определим, какие натуральные числа могут быть вообще воспроизведены в виде суммы двух квадратных чисел. Ставится вопрос: при каких значениях n может быть решено уравнение (2). Предположим, что некоторое неделимое число p является делителем числа n. Если выполняется условие

$p \mid x$, то, видимо, выполняется и условие $p \mid y$, а из последнего следует выполнение соотношения $p^2 \mid n$. Тогда при делении уравнения на p^2 получится уравнение подобного типа. Итак, можно предполагать правильность соотношения $p \nmid x$. В таком случае из уравнения (2) получится конгруэнтность типа $y^2 \equiv -x^2 \pmod p$. На основе соотношения $(p, x) \equiv 1$ существует такое значение u, в отношении которого имеет место конгруэнтность $ux \equiv 1 \pmod p$. Из этих конгруэнтностей получится новая конгруэнтность $(uy)^2 \equiv -1 \pmod p$; следовательно, неделимое число p должно представить собой такое число, в отношении которого число -1 является квадратичным остатком. А такое условие выполняется, как было показано выше, для неделимых чисел вида $4k + 1$. Получится, что если в разложении числа n на произведение неделимых чисел присутствует некоторое неделимое число вида $4k - 1$, причем в степени нечетного числа, то число n не может быть воспроизведено по уравнению (2). Предположим, что в разложении числа n на произведение неделимых чисел каждое неделимое число вида $4k - 1$ присутствует в степени четного числа. Может быть доказано, что если p представляет собой неделимое число вида $4k + 1$, то оно не только является квадратичным остатком типа -1 по модулю p, но и может быть воспроизведено в виде согласно уравнению (2). Если один из членов является 0, то некоторое квадратное число может быть воспроизведено в такой же форме. Каждое из возможных чисел n может быть записано в виде произведения некоторого квадратного числа и определенного количества неделимых чисел вида $4k + 1$, т. е. в виде произведения таких чисел, каждое из которых может быть записано в виде суммы двух квадратных чисел. Но в таком случае и произведение может быть воспроизведено таким же образом, что доказывается следующим простым тождеством:

$$(x^2 + y^2)(u^2 + v^2) = x^2u^2 + y^2u^2 + x^2v^2 + y^2v^2 = (x^2u^2 + 2xyuv + \\ + y^2v^2) + (y^2u^2 - 2xyuv + x^2v^2) = (xu + yv)^2 + (yu - xv)^2.$$

В итоге оказывается действительой следующая теорема:

Уравнение (2) может быть решено тогда и только тогда, если в разложении числа n на произведение неделимых чисел ни одно из неделимых чисел вида $4k - 1$ не присутствует в степени нечетного числа.

Аналогично можно исследовать вопрос о воспроизведении чисел в виде суммы нескольких квадратных чисел. В виде суммы трех квадратных чисел не могут быть представлены те и только те числа, которые имеют вид $4^k(8n + 7)$, где k — положительное целое число и n — неотрицательное целое число.

В отношении суммы из четырех чисел действительна *теорема Лагранжа: каждое натуральное число может быть воспроизведено в виде суммы четырех неотрицательных целых чисел.*

Исследование гауссовых целых чисел. Здесь также возникает проблема определения количества решений исследуемого уравнения. В отношении такой проблемы ответы могут быть найдены с помощью исследования так называемых *гауссовых целых чисел* или теории чисел так называемых *целых кватернионов*. Гауссовы целые числа представляют собой комплексные числа вида $a + bi$, где a и b — целые числа. Данные числа связаны с уравнением (2) через разложение вида $x^2 + y^2 = (x + yi)(x - yi)$. В данном разделе математики могут быть достигнуты значительные результаты благодаря применению гауссовых целых чисел, так как удается преобразовать «аддитивную проблему» в «мультипликативную». Кватернионы подобны комплексным числам, но они представляют собой не двухчленные, а четырехчленные выражения. Проблемы, связанные с кватернионами, более подробно рассматриваются при обсуждении проблем алгебры.

Исследование уравнений типа Пелл. В области представления квадратичных разложений до сих пор рассматривались такие уравнения, в которых все коэффициенты были 1. Ниже рассматривается проблема так называемого *уравнения Пелла:*

$$(3) \qquad x^2 - dy^2 = 1, \qquad \text{где } d \text{ — целое число.}$$

При любом значении d настоящее уравнение имеет корни $x = 1$, $y = 0$ и $x = -1$, $y = 0$. Эти решения называются *тривиальными решениями*. В случае $d = 0$ y может принимать любое значение, в то время как или $x = 1$, или $x = -1$. Если d представляет собой отрицательное или положительное квадратное число, то данное уравнение имеет только тривиальные решения (и, кроме того, при $d = -1$ существуют еще два решения). Приведенные здесь частные случаи не усложняют обсуждение данной тематики; исследование уравнения вышеприведенного типа представляет собой интерес в тех случаях, когда d — положительное число, не являющееся квадратным числом.

Если буквы x, y и u, v обозначают два (не обязательно различных) решения уравнения (3), то выполняются соотношения вида:

$$(x + y\sqrt{d})(x - y\sqrt{d}) = x^2 - y^2d = 1 \qquad \text{и}$$
$$(u + v\sqrt{d})(u - v\sqrt{d}) = u^2 - dv^2 = 1.$$

Применением соотношения $(x + y\sqrt{d})(u + v\sqrt{d}) = (xu + \partial yv) + (xv + yu)\sqrt{d}$ может быть доказано, что выражение $(xu + \partial yv)$, $(xv + yu)$ также является решением уравнения (3). Если ни одно из заданных решений не является тривиальным, то полученное в последнем случае решение отличается от каждого предыдущего. Можно показать, что если d — положительное число, не являющееся квадратным, то существует нетривиальное решение уравнения (3). Очевидно, среди нетривиальных решений можно обнаружить такое положительное решение x_1, y_1, в отношении которого выражение $|x_1 + y_1\sqrt{d}|$ примет минимальное значение. Выражения $x_1 + y_1\sqrt{d}$ в положительной целой степени однозначно могут быть записаны в виде $(x_1 + y_2\sqrt{d})^n = x_n + y_n\sqrt{d}$, где x_n и y_n — целые числа.

Можно доказать, что все решения уравнения (3) записываются в виде:

$$x_n, y_n; \quad x_n, -y_n; \quad -x_n, y_n; \quad -x_n, -y_n; \quad (n = 1, 2, \ldots)\ 1, 0; \quad -1, 0.$$

Например, в случае $d = 2$ одним из решений является $x = 3$, $y = 2$, из чего определяются два новых решения: $(3 + 2\sqrt{2})^2 = 17 + 12\sqrt{2}$ (действительно, $17^2 - 2 \cdot 12^2 = 289 - 288 = 1$); и $(3 + 2\sqrt{2})^3 = 99 + 70\sqrt{2}$ (действительно, $99^2 - 2 \cdot 70^2 = 9801 - 9800 = 1$).

РАЗЛОЖЕНИЯ ЧИСЕЛ БОЛЕЕ ВЫСОКОГО ПОРЯДКА

Большая теорема Ферма. Рассматриваются два разложения такого типа. Первым из них является так называемая *большая теорема Ферма»* или, точнее, *предположение Ферма*:

Если n — целое число больше 2, то уравнение

$$x^n + y^n = z^n$$

имеет только тривиальное решение при целых значениях x, y и z. Настоящее утверждение можно доказать для случая, когда n представляет собой неделимое число и если $n = 4$. Утверждение доказано для различных случаев; например, утверждение правильно при $n \leq 100$.

Проблема Уоринга является второй проблемой настоящей главы. Как мы уже знаем, каждое натуральное число может быть записано в виде суммы четырех квадратных чисел. Гильбертом было доказано, что если n представляет собой натуральное число, то существует такое натуральное число k, зависящее от числа n, что каждое натуральное число может быть воспроизведено в виде суммы n-ых степеней чисел, количество которых составляет именно k.

АЛГЕБРАИЧЕСКИЕ И ТРАНСЦЕНДЕНТНЫЕ ЧИСЛА

ПОНЯТИЕ АЛГЕБРАИЧЕСКИХ И ТРАНСЦЕНДЕНТНЫХ ЧИСЕЛ

Исследования, проводимые в области теории чисел, вышли из рамок области рациональных чисел именно на основе результатов Куммера. Но иррациональные числа характеризуются весьма противоречивыми свойствами, их делят на два класса: на алгебраические и трансцендентные.

Некоторое число α называется алгебраическим числом, если существует такой многочлен $a_0 + a_1x + a_2x^2 + \ldots + a_{n-1}x^{n-1} + a_nx^n$ с рациональными коэффициентами, корнем которого является α. Если коэффициенты этого многочлена представяют собой целые числа, то α называется алгебраическим целым числом.

Если нельзя найти такой многочлен для числа α, то α представляет собой трансцендентное число.

Если для числа α может быть найден хотя бы один многочлен такого типа порядка n, корнем которого является α, но уже нельзя найти многочлен более низкого порядка, то число α называется алгебраическим целым числом порядка n.

Очевидно, алгебраическими числами первого порядка являются именно рациональные числа, и алгебраические целые числа первого порядка представляют собой именно рациональные целые числа.

Например, число $\sqrt{d}$ при уравнении типа Пелл представляет собой алгебраическое целое число второго порядка, если d — не квадратное число, так как оно является корнем многочлена $x^2 - d$, но не является корнем ни одного многочлена первого порядка. Может быть доказано, что для любого натурального числа n можно найти алгебраическое число n-го порядка; таким числом является, например, число $\sqrt[n]{2}$.

ДИОФАНТОВА АППРОКСИМАЦИЯ

Нетрудно проводить различые операции дробными числами, но труднее манипулировать иррациональными числами. Поэтому целесообразно *аппроксимировать* — реализовать приближение иррациональных чисел с помощью рациональных чисел.

Для исследования аппроксимации иррациональных чисел с помощью рациональных чисел рассматривается следующая проблема. Пусть α представляет собой иррациональное число. Рассматривается положительная сторона численной прямой как «нить, являющаяся бесконечной справа»:

$$\alpha,\ 2\alpha,\ \ldots,\ n\alpha,\ (n+1)\alpha.$$

Пустим данную нить по кругу, *длина окружности* которого составляет 1. Отмеренные числа расположены в $n+1$ точках по окружности круга. Пусть отправляемся из одной этих точек в одно направление и посмотрим следующую в данном направлении точку для всех точек. Не может быть, чтобы во всех случаях дистанция между соседними точками, *отмеренная по окружности круга*, была по меньшей мере $\frac{1}{n}$. В таком случае, чтобы прийти в исходную точку, следовало бы пройти дистанцию $\frac{n+1}{n}$ по окружности круга, исходя из первой точки. Такая дистанция была бы больше 1, в то время как пройдя по окружности до исходной точки, можно преодолеть только такую дистанцию, которая соответствует длине окружности круга, составляющей именно 1. Таким образом, имеются две точки на окружности круга — предположим, точки $u\alpha$ и $v\alpha$ — дистанция между которыми меньше $\frac{1}{n}$. Поскольку по окружности круга совпадают те точки численной прямой, разность между которыми представляет собой целое число, то вышеприведенный результат означает, что путем вычитания соответствующего целого числа из числа $u\alpha$ разность между полученным таким образом числом и числом $v\alpha$ будет меньше $\frac{1}{n}$. Т. е. существует такое целое число p, при котором выполняется условие $|(u-v)\alpha - p| < \frac{1}{n}$. Можно предполагать правильность соотношения $u > v$, и, поскольку значение числа v по меньшей мере составляет 1, то положительное число q будет: $q = u - v \leq n$. Если поделить полученное неравенство на q, то получится следующий результат:

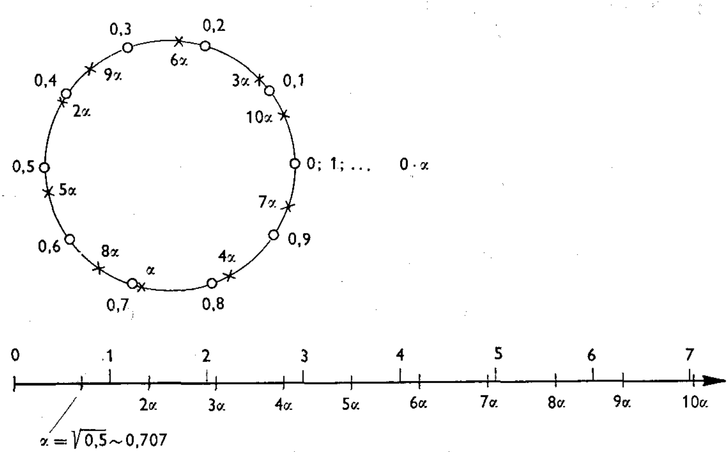
Если α — иррациональное число и n — любое натуральное число, то существует такое целое число p и такое натуральное число q, которое не больше чем n, в отношении которых выполняется условие:

(1) $$\left|\alpha - \frac{p}{q}\right| < \frac{1}{nq}.$$

Из этого условия непосредственно получится неравенство:

(2) $$\left|\alpha - \frac{p}{q}\right| < \frac{1}{q^2}$$

На рисунке представлен случай, когда $\alpha = \sqrt{0{,}5}$ и $n = 10$. Чтобы рисунок стал более наглядным, круг с единичной длиной окружности увеличен в 5 раз. Из рисунка ясно, что при любой из пар 10α и 3α; 9α и 2α; 8α и α; 7α и $0 = 0\alpha$ дистанция между точками меньше 0,1. На числовой оси видно, что к числу 5 ближе всех расположена точка 7α. Таким образом, имеет место неравенство $|7\alpha - 5| < \frac{1}{10}$, т. е. неравенство $\left|\alpha - \frac{5}{7}\right| < \frac{1}{70}$.



Действительно, $70\left(\frac{5}{7} - \sqrt{0{,}5}\right) = 50 - 70\sqrt{0{,}5} < 50 - 49{,}4 = 0{,}6 < 1$, и на основе настоящего неравенства выполняется неравенство, представленное выше.

Покажем, что *бесконечное количество дробей вида* $\frac{p}{q}$ *характеризуется свойством* (2).

Предположим, что существует только ограниченное количество таких дробей. Пусть Q является общим знаменателем, дробь $\frac{P}{Q}$ имеет значение, меньшее α, но дробь $\frac{P+1}{Q}$ — значение, больше α. (Из-за иррациональности числа α равенство исключается.) Пусть δ представляет собой такое число, которое совпадает с более маленьким из чисел $\alpha - \frac{P}{Q}$ и $\frac{P+1}{Q} - \alpha$. Поскольку δ является положительным, то существует такое натуральное число n, в отношении которого выполняется неравенство $\frac{1}{n} < \delta$. Но в таком случае любая другая дробь со знаменателем Q отличается от значения α по меньшей мере на величину $\frac{1}{n}$, т. е. если выражение $\frac{a}{b}$ представляет собой любую из дробей, рассмотренных в исходном случае, то имеет место соотношение :

$$(3) \qquad \left|\alpha - \frac{a}{b}\right| > \frac{1}{n}.$$

Если теперь повторить рассуждение, приведенное выше в отношении настоящего натурального числа n, то получится результат, по которому существует така [дробь $\frac{p}{q}$, для которой неравенство

$$(4) \qquad \left|\alpha - \frac{p}{q}\right| < \frac{1}{nq} < \frac{1}{n}$$

выполняется, в то время, как настоящее соотношение не будет действительно для дробей, заданных на исходный случай ;. получено противоречие в исходном предположении.

Итак, при любом иррациональном числе можно найти бесконечное количество таких дробей, разность которых от данного числа меньше, чем квадрат знаменателя. Настоящее утверждение не относится к рациональным числам. Поедположим, выполняется неравенство $\left|\frac{a}{b} - \frac{p}{q}\right| < \frac{1}{q^2}$. Поскольку значение величины q составляет, по меньшей мере, 1, то значение дроби $\frac{p}{q}$ расположено между значениями $\frac{a}{b} - 1$ и $\frac{a}{b} + 1$. С другой стороны, имеет место соотношение: $\frac{a}{b} - \frac{p}{q} = \frac{aq - bp}{bq}$. Если две данные дроби различаются, то абсолютное значение данной дроби со знаменателем

bq составляет, по меньшей мере $\frac{1}{bq}$. Если сравнить полученный результат с исходным предположением, то получится неравенство $\frac{1}{bq} < \frac{1}{q^2}$, т. е. $q < b$. Но между значениями $\frac{a}{b} - 1$ и $\frac{a}{b} + 1$ имеется ограниченное количество дробей, знаменатель которых меньше b.

Воэникает вопрос, можно ли аппроксимировать иррациональные числа «еще лучше». Докажем, что нет. Например, в отношении числа $\sqrt{2}$ существует ограниченное количество дробей вида $\frac{p}{q}$, для которых выполняется неравенство $\left|\sqrt{2} - \frac{p}{q}\right| < \frac{1}{q^3}$. Значение некоторой такой дроби должно находиться между значениями $\sqrt{2} - 1$ и $\sqrt{2} + 1$. С другой стороны, имеет место неравенство $\left|\sqrt{2} + \frac{p}{q}\right| \leq \left|\sqrt{2} - \frac{p}{q}\right| + 2\sqrt{2} < 1 + 2\sqrt{2}$. Если умножить данное неравенство на исходное, то получится новое неравенство $\left|2 - \frac{p^2}{q^2}\right| < \frac{1 + 2\sqrt{2}}{q^3}$. Значение дроби, стоящей в левой стороне последнего неравенства, отличается от 0 на основе иррациональности числа $\sqrt{2}$, и поэтому оно составляет, по меньшей мере, $\frac{1}{2}$. Таким образом, действительно и соотношение $\frac{1}{q^2} < \frac{1 + 2\sqrt{2}}{q^3}$, т. е. $q < 1 + 2\sqrt{2}$. Но между значениями $\sqrt{2} - 1$ и $\sqrt{2} + 1$ находится ограниченное количество дробей, знаменатель которых — (в соответствии с доказанным утверждением) меньше $1 + 2\sqrt{2}$.

Путем небольшой модификации приведенного метода подобный вывод может быть сделан вообще для алгебраических чисел:

Теорема Льювилля: *Если α представляет собой алгебраическое число порядка n, то существует ограниченное количество дробей вида $\frac{p}{q}$, для которых выполняется неравенство*

$$\left|\alpha - \frac{p}{q}\right| < \frac{1}{q^{n+1}}.$$

ИЗВЕСТНЫЕ ТРАНСЦЕНДЕНТНЫЕ ЧИСЛА

С помощью теоремы Льювилля можно воспроизводить трансцендентные числа. Впрочем, именно Льювилль определил так называемые *числа Льювилля*, понятие которых представлено ниже. В ходе истории математики это были первыми заданными трансцендентными числами и с помощью этих чисел удалось доказать существование трансцендентных чисел.

По теореме Льювилля, если для некоторого числа α при *любом* натуральном числе n можно найти бесконечное количество таких дробей, разность которых от этого числа меньше, чем обратное значение $(n+1)$, то данное число не может представлять собой алгебраическое число порядка n при никаком значении n. Следовательно, оно дольжно быть трансцендентным числом. Числа Льювилля задаются в виде бесконечной прогрессии, сходимость которой доказывается с помощью метода, представленного в главе «Анализ». Пусть число α записывается в виде

$$\alpha = \frac{1}{2} + \frac{1}{2^2} + \frac{1}{2^6} + \frac{1}{2^{4!}} + \ldots + \frac{1}{2^{k!}} + \ldots$$

Сумма первых k членов представляет собой рациональное число. Пусть она обозначается (в упрощенном виде) как $\frac{p_k}{q_k}$. (Очевидно, соотношение $q_k = 2^{k!}$ является правильным.) О сумме остальных членов можно сказать, что она составляет, по крайней мере, $\frac{2}{2^{(k+1)!}} = \left(\frac{1}{2^{k!}}\right)^k$. Следовательно, если выполняется условие $k > n$, то имеет место соотношение $\left|\alpha - \frac{p_k}{q_k}\right| < \frac{1}{(q_k)^k} \leq$ $\leq \frac{1}{(q_k)^{n+1}}$. Поскольку существует бесконечное количество чисел k, больших n, может быть найдено бесконечное количество соответствующих дробей, т. е. число α — действительно трансцендентно. Очевидно, что таким же образом будет получено трансцендентное число и в том случае, если в знаменателях записываются соответствующие степени некоторого числа вместо 2, которое больше 1. Такие числа называются *числами Льювилля*.

На этой основе Гермитом было доказано, что e (основное число натуральных логарифмов) является трансцендентным, а позднее Линдеманном была доказана трансцендентность и числа π ($= 3{,}14\ldots$).

ГЕОМЕТРИЧЕСКАЯ ТЕОРИЯ ЧИСЕЛ

ТЕОРЕМА МИНКОВСКОГО

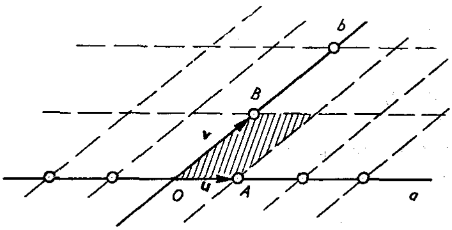
Геометрической теорией чисел называется такая область теории чисел, в которой теоремы по теории чисел доказываются но основе применения геометрических методов.

В области геометрической теории чисел понятие *сетки параллелограммов* играет центральную роль.

Сетка параллелограммов. Возьмем систему координат в плоскости и начертим две (различных) прямых a и b, проходящих через начало координат. Затем измерим некоторую фиксированную дистанцию на прямой a, начиная с начала координат, по обоим направлениям. Такой же процесс повторим и на прямой b; здесь фиксированная дистанция может отличаться от замеренной на прямой a. Начертим параллельные с прямой b линии через точки, полученные на прямой a, а затем — параллельные с прямой a линии через точки, полученные на прямой b. В таком случае любая прямая, параллельная прямой a, будет пересекать все прямые, параллельные прямой b.

Совокупность полученных таким образом точек пересечения называется сеткой параллелограммов, отдельные точки пересечения — сетевыми точками, а прямая, связывающая две сетевых точки — сетевой прямой.

Решетчатый параллелограмм. Пусть соседняя с началом координат точка на прямой a обозначается через A, а на прямой b — через B, вектор, направленный от начала координат в сторону точки A — через $\mathbf{u}$, а в сторону B — через $\mathbf{v}$. Очевидно, что некоторая точка представляет собой сетевую точку тогда и только тогда, если в отношении вектора $\alpha\mathbf{u} + \beta\mathbf{v}$, направленного от начала координат в сторону данной точки, числа α и β представляют собой целые числа.



Параллелограмм, двумя соседними сторонами которого являются векторы $\mathbf{u}$ и $\mathbf{v}$, называется основным параллелограммом сетки. Если четыреугольник, определенный четырьмя сетевыми точками, является параллелограммом, то он называется решетчатым параллелограммом. Если внутри некоторого решетчатого параллелограмма нет ни одной сетевой точки, то данный параллелограмм называется пустым.

Можно показать, что площадь любых двух пустых решетчатых параллелограммов одинакова и, видимо, равна площади основного параллелограмма, так как он также является решетчатым.

Пусть задана некоторая сетка из параллелограммов, в которой площадью основного параллелограмма является величина Δ.

Рассматривается некоторая область

выпуклая,

центрально симметричная по отношению к началу координат,

замкнутая,

площадь которой не больше 4Δ.

(Соответствующие понятия приведены в разделе «Геометрия».)

В этом случае *по теореме Минковского данная область содержит сетевую точку, отличающуюся от начала координат.*

Рассмотрим прямую a на заданной сетке и каждую вторую прямую из параллельных с прямой a линий; аналогично, прямую b и каждую вторую из прямых, параллельных с ней. Если «разрезать» данную сетку по этим прямым, то вся плоскость распадется на такие решетчатые параллелограммы, каждый из которых содержит по 4 равных с основным параллелограммом решетчатых параллелограмма. Их площадь составляет 4Δ. Сдвинем данные параллелограммы таким образом, чтобы одна из их вершин была в начале координат.

Рассмотрим один из таких параллелограммов: так как была выбрана только каждая вторая прямая, то вектор, направленный в любую вершину, имеет вид $2\alpha\mathbf{u} + 2\beta\mathbf{v}$, где α и β — целые числа. Предположим, что данный вектор направлен в сторону именно той вершины параллелограмма, которая попала в точку начала координат после сдвига. Если некоторая точка P настоящего параллелограмма перемещается в точку P_1 после сдвига, то вектор, направленный из точки P_1 в сторону точки P, очевидно, также записывается в виде $2\alpha\mathbf{u} + 2\beta\mathbf{v}$. Представим себе, что сетка начерчена на прозрачной бумаге и вышеприведенная область зачернена на бумаге. При разрезании сетки выбранная область также будет разрезана и при осуществлении сдвига разрезанные части переместятся в параллелограмм площадью в $4\varDelta$. Поскольку площадь выбранной области больше $4\varDelta$, то по меньшей мере две из этих частей где-то совпадут друг с другом. Пусть Q представляет собой точку, характеризуемую таким свойством. Поскольку здесь по меньшей мере две части области совпадают друг с другом, то в ходе сдвига по меньшей мере две точки исходной области переместятся в точку Q. Пусть эти две точки обозначаются знаками Q_1 и Q_2. Это означает, что векторы, направленные из точки Q в сторону точек Q_1 и Q_2, могут быть записаны в виде $2\gamma_1\mathbf{u} + 2\delta_1\mathbf{v}$ и $2\gamma_2\mathbf{u} + 2\delta_2\mathbf{v}$, где величины γ_1, γ_2, δ_1, δ_2 — целые числа. Если $\mathbf{q}$ является вектором, направленным из начала координат в точку Q, то вектор, направленный из начала координат

в точку Q_1, записывается в виде $2\gamma_1\mathbf{u} + 2\delta_1\mathbf{v} + \mathbf{q}$,
а в точку Q_2 — в виде $2\gamma_2\mathbf{u} + 2\delta_2\mathbf{v} + \mathbf{q}$.

До сих пор имелись условия, представляющие ограничение в отношении площади исследуемой области (замкнутость области также была использована). Теперь будут применены свойства области по центральной симметричности и выпуклости. Точки Q_1 и Q_2 были выбраны таким образом, что обе являлись точками выбранной области. Пусть R обозначает отображение точки Q_2 по началу координат, S — точку, разделяющую пополам отрезок Q_1R. Точка R принадлежит к области, так как она является центральной симметричной точкой по отношению к началу координат. В этом случае на основе выпуклости области точка S также принадлежит к данной области и представляет собой сетевую точку. С учетом свойства отображения, вектор, направленный из начала координат в точку R, описывается выражением $-2\gamma_2\mathbf{u} - 2\delta_2\mathbf{v} - \mathbf{q}$. Вектор, направленный из начала координат в точку, разделяющую пополам дистанцию Q_1R, может быть получен таким образом, что берется половина суммы векторов, направленных из начала координат в конечные точки. Поэтому вектор, направленный из начала координат в точку S, описывается в виде $(\gamma_1 - \gamma_2)\mathbf{u} + (\delta_1 - \delta_2)\mathbf{v}$. Приведенные в данном выражении все числа являются целыми, их разность — также целое. В таком виде могут быть записаны только векторы, направленные в сетевые точки, т. е. S представляет собой сетевую точку. Необходимо убедиться и в том, что точка S отличается от точки начала координат. Но это очевидно, так как если бы точка S представляла собой начало координат, то отображением точки R по началу координат была бы точка Q_1, и точки Q_1 и Q_2 равнялись бы между собой.

Рассмотрим одно применение теоремы Минковского в области теории чисел. Предположим, что если p представляет собой неделимое число вида $4k+1$, то существует такое число a, в отношении которого выполняется условие $a^2 \equiv -1 \pmod{p}$. Из этого соотношения будет доказано, что число p может быть записано в виде x^2+y^2, где x и y — натуральные числа. Рассматриваются такие точки плоскости, в отношении координат x и y которых имеют место виды записи:

$$x = \alpha p + \beta a, \qquad y = \beta; \qquad \text{где } \alpha \text{ и } \beta \text{ — целые числа.}$$

Если $\mathbf{u}$ обозначает вектор, направленный из начала координат в точку $(p;\ 0)$ и $\mathbf{v}$ — вектор, направленный в точку $(a;\ 1)$, то вектор, направленный в точку $(x;\ y)$ записывается в виде $\alpha\mathbf{u}+\beta\mathbf{v}$, и данные точки образуют некоторую из параллелограммов. Четырьмя вершинами основного параллелограмма данной сетки являются точки $(0;\ 0)$, $(p;\ 0)$, $(a;\ 1)$ и $(a+p;\ 1)$. Настоящий параллелограмм лежит на оси x, длина его основы — p; его высота — 1 и, таким образом, его площадь составляет $\Delta = p$. Пусть вычисляется квадрат дистанции некоторой точки настоящей сетки от начала координат:

$$x^2+y^2 = \alpha^2p^2 + 2\alpha\beta\, pa + \beta^2a^2 + \beta^2 = p(\alpha^2p + 2\alpha\beta a) + \beta^2(a^2+1)\,.$$

Поскольку предполагалась правильность соотношения $a^2 \equiv -1 \pmod{p}$, то и второй член в правой стороне данного соотношения является делимым на p. Иными словами, для точек данной сетки выполняется условие:

$$x^2+y^2 = \gamma p, \qquad \text{где } \gamma \text{ — целое число.} \tag{1}$$

Рассматривается круг, центром которого является начало координат и радиусом — $2\sqrt{\dfrac{p}{3}}$.

Данный круг, очевидно, выпуклый, замкнутый и центрально симметричный по началу координат. Площадь круга составляет $\dfrac{4p\pi}{3} > 4p = 4\Delta$.

Таким образом, можно применить теорему Минковского; следовательно, данный круг содержит сетевую точку, отличающуюся от начала координат. В отношении координат $(x;\ y)$ данной точки выражение x^2+y^2 является положительным, так как данная точка отличается от начала координат, и выражение x^2+y^2 меньше, чем квадрат радиуса круга, поскольку данная точка расположена внутри круга. Таким образом, соотношение

$$0 < x^2+y^2 \leq \frac{4p}{3} < 2p \tag{2}$$

является правильным. Если сравнить неравенство (1) с неравенством (2), оказывается:

$$0 < \gamma p < 2p, \text{ т. е. если поделить на } p\text{: } 0 < \gamma < 2.$$

Учитывая, что γ представляет собой целое число, единственным решением данного неравенства может быть только $\gamma = 1$, откуда для координат рассматриваемой точки $(x; y)$ из неравенства (1) получится соотношение:

$$x^2 + y^2 = p.$$

Поскольку числа x и y — целые, то неделимое число p воспроизведено в желаемом виде. (Знак не учитывается, поскольку была проведена операция возведения в квадрат.)

КРАТКИЙ ОБЗОР ИСТОРИИ ТЕОРИИ ЧИСЕЛ

При исследовании операций, проводимых с числами, рождаются идеи, относящиеся к проблемам теории чисел. Такие идеи появлялись у ученых Еги пта, а затем в работах Пифагора (принцип абакуса). Были введены понятия квадратных чисел, кубовых чисел, тройных чисел и совершенных чисел (равняющихся сумме их делителей), так называемых дружеских чисел (таких пар чисел, каждое из которых равняется сумме делителей другого числа пары) и т. д. В связи с этими проблемами древние греки решали и сложные задачи.

В этой области исследований среди греческих математиков выделяется Евклид. Метод определения наибольшего общего делителя до сих пор называется *евклидовым алгоритмом*. Впервые он доказал теорему, по которой существует бесконечное количество неделимых чисел. Одним из самых древних математиков считается Эратостен (276—195 до н. э.) (по существу, его считали астрономом). Его именем назван метод, по которому могут быть определены неделимые числа, меньшие некоторого заданного числа (сито Эратостена).

В работах Диофанта (325—409) можно найти некоторые корни проблем, решенних существенно позже: им была задана теорема, по которой каждое неквадратное число может быть воспроизведено в виде суммы по крайней мере четырех квадратных чисел.

Индийский математик Айа-Батха (род. ок. 476 г. н. э.) определил метод вычисления суммы квадратов и кубов чисел, следующих друг за другом. В результате деятельности арабского математика аль-Хорезми (около 820 г. н. э.) была распространена десятичная система чисел. В работах Алькайами впервые было найдено, что нерешительное уравнение вида $x^3 + y^3 = z^3$ не может иметь решения из целых чисел; этим он как бы опередил Ферма. В работах Ибн Абанна (примерно в 1222 г.) проводились исследования по суммам вида $1^p + 2^p + 3^p + \ldots n^p$ $(p = 1, 2, 3, 4)$.

Фибоначчи (родился в 1180 г.) рассматривал решения из целых чисел уравнения вида $x^2 + y^2 = 2z^2$. Его именем названа последовательность ($a_1 = 0$, $a_2 = 1$, $a_k = a_{k-1} + a_{k-2}$.) Тарталья (1500—1577) представил результаты, подобные треугольнику Паскаля.

В века бурного развития математики невозможно привести имена всех ученых, которые внесли вклад в построение теории чисел. Почти все известные математики активно занимались этой проблемой, кроме того, и много «любителей» достигли результатов. Здесь приводятся только самые известные специалисты теории чисел.

П. Ферма (1601—1665): его деятельность в области теории чисел считается самой известной. Им было оформлено много теорем, доказательства которых или не опубликованы, или потеряны. Долгое время интерес многих математиков был привлечен проблемой нахождения этих доказательств.

Впервые Б. Паскалем (1623—1662) был применен метод полной индукции.

Крупным специалистом в области теории чисел являлся Л. Эйлер (1707—1783): впервые он представил доказательство теоремы Ферма в 1736 г; ввел понятие функции теории чисел $\varphi(n)$ и исследовал свойства данной функции; с ее помощью ему удалось обобщить утверждение Ферма. В этой же области работал Сэр Джон Уильсон (1741—1793).

К. Ф. Гаусс (1777—1855) также занимался проблемами теории чисел, особенно в области квадратичных видов и теории конгруэнтностей. Последняя теория создана Гауссом.

В деятельности А. М. Лежандра (1752—1833) теория чисел также играла доминирующую роль. Он представил доказательство теоремы Ферма для частного случая. Теорема квадратичного обращения до сих пор носит название теоремы Лежандра.

П. Г. Лежен-Дирихле (1805—1859) в возрасте 20 лет доказал теорему Ферма для случая, когда $n = 5$ а затем для случая $n = 14$. Представлением этих и многих других важных теорем его деятельность обогатила теорию чисел.

Б. Риман (1826—1866) ввел понятия функции $\zeta(s) = \sum \frac{1}{n^s}$.

Ж. Бертра (1822—1900) представил утверждение, по которому между числами $a > 1$ и $2a$ всегда может быть обнаружено неделимое число.

В дальнейщем наиболее выдающимися деятелями в области теории чисел явились: Г. Жюлия (1893—), П. Гумберт (1859—1921), Ж. Адамар (1865—1963) Шевалье (1909—) и другие.

Выдающимся представителем советской школы в области теории чисел является И. М. Виноградов (1891—).

В Венгрии наиболее результативной является деятельность П. Эрдёша (1913—), П. Турана (1911—) и А. Рени (1921—1970).

ГЕОМЕТРИЯ

ОСНОВНЫЕ ПОНЯТИЯ

ПРОИСХОЖДЕНИЕ ГЕОМЕТРИЧЕСКИХ ПОНЯТИЙ

Геометрия — одна из наиболее древних и ранее других систематизированная ветвь математики. Она возникла из задач, связанных с различными измерениями земельных участков. Основные предложения и теоремы геометрии произошли из опыта, накопленного человечеством в течение многих веков. Простейшие геометрические понятия появились в результате постепенного и очень длительного абстрагирования. В процессе абстрагирования отвлекаются от большинства материальных свойств предметов и рассматривают, например, только их форму. Так можно прийти к понятиям тела, поверхности, кривой линии и точки. Таким образом, геометрические понятия происходят из окружающей нас действительности и упрощенно отражают определенные соотношения, имеющие место в материальном мире.

МЕТОД ПОСТРОЕНИЯ ГЕОМЕТРИИ, ОСНОВНЫЕ ПОНЯТИЯ, АКСИОМЫ

Метод построения геометрии как науки сформировался уже более двух тысяч лет назад. Суть этого метода в следующем: некоторые *основные понятия* считаются известными из наглядных соображений, их не объясняют и не определяют с помощью других понятий. Таковы, например, понятия точки, прямой, плоскости, понятие принадлежности, а также понятие равенства отрезков или углов. Эти понятия используют в дальнейшем при введении новых, таких, например, как полупрямая, отрезок, угол и тому подобное.

После этого высказывают некоторые основные предложения, *аксиомы*, которые считаются справедливыми также из наглядных соображений. Из них уже логическим путем строится вся геометрия.

Такой метод построения геометрии впервые применил Евклид (около 325 года до н. э.) в своем труде, который он назвал «Начала». Первое обоснование этого метода, отвечающее более современным требованиям, связано с именем Д. Гильберта (1899). По существу, Гильберт построил геометрию, основываясь на аксиомах, которые мы приводим ниже (для простоты мы используем в этих аксиомах уже введенные понятия).

I. АКСИОМЫ ПРИНАДЛЕЖНОСТИ

1. Любым двум различным точкам A и B всегда соответствует некоторая прямая a, проходящая через эти точки.

2. Любым двум различным точкам A и B соответствует не более одной прямой, проходящей через эти точки.

3. Всякой прямой принадлежат по крайней мере две точки. Существуют три точки, не лежащие на одной прямой.

4. Любым трем точкам A, B и C, не лежащим на одной прямой, соответствует по крайней мере одна плоскость α, проходящая через эти точки. Всякой плоскости принадлежит по крайней мере одна точка.

5. Любым трем точкам A, B и C, не лежащим на одной прямой, соответствует не более одной плоскости, проходящей через эти точки.

6. Если две различные точки A и B прямой a принадлежат некоторой плоскости α, то этой плоскости принадлежат все точки прямой a.

7. Если две плоскости α и β имеют общую точку A, то они имеют еще по крайней мере одну общую точку B, отличную от A.

8. Существуют по крайней мере четыре точки, не принадлежащие одной плоскости.

II. АКСИОМЫ ПОРЯДКА

1. Если точка B лежит между точками A и C, то A, B и C являются различными точками некоторой прямой, причем можно сказать также, что точка B лежит между C и A.

2. Для различных точек A и C на прямой AC найдется по крайней мере одна точка B такая, что C лежит между A и B.

3. Из любых трех точек одной прямой одна и только одна лежит между двумя другими.

4. Пусть A, B, C — три точки, не лежащие на одной прямой, и a — некоторая прямая на плоскости ABC, не проходящая ни через одну из этих точек. Тогда если прямая a пересекает отрезок AB, то она непременно пересекает также или отрезок BC, или AC.

III. АКСИОМЫ РАВЕНСТВА (СОВПАДЕНИЯ)

1. Пусть A и B — две различные точки прямой a, и A' — точка прямой a'. Тогда на одной из полупрямых, определенных заданием точки A' на прямой a', всегда найдется такая точка B', что отрезок AB при наложении совпадает с отрезком $A'B'$, то есть эти отрезки равны между собой. Это обозначается так:

$$AB \equiv A'B'.$$

2. Если каждый из отрезков $A'B'$ и $A''B''$ равен отрезку AB, то отрезок $A'B'$ равен отрезку $A''B''$.

3. Пусть на прямой a заданы два отрезка, AB и BC, не имеющие общих точек, и пусть на той же прямой или на некоторой другой прямой a' заданы отрезки $A'B'$ и $B'C'$, также не имеющие общих точек. Тогда если $AB \equiv A'B'$ и $BC \equiv B'C'$, то $AC \equiv A'C'$.

4. Пусть на плоскости заданы $\measuredangle (h, k)$ (угол между лучами (полупрямыми) h и k), прямая a' и одна из соответствующих ей полуплоскостей.

Если h' обозначает одну из полупрямых прямой a', то существует одна и только одна полупрямая k' такая, что $\sphericalangle(h, k)$ при наложении совпадает с $\sphericalangle(h', k')$, то есть равен ему:

$$\sphericalangle(h, k) \equiv \sphericalangle(h', k'),$$

причем все внутренние точки угла $\sphericalangle(h', k')$ лежат в заданной полуплоскости.

Всякий угол равен самому себе, то есть всегда имеет место равенство

$$\sphericalangle(h, k) \equiv \sphericalangle(h, k).$$

5. Если для треугольников ABC и $A'\ B'\ C'$ выполняются равенства

$$AB \equiv A'B', \quad AC \equiv A'C' \quad \text{и} \quad \sphericalangle BAC \equiv \sphericalangle B'A'C',$$

то справедливо и равенство

$$\sphericalangle ABC \equiv \sphericalangle A'B'C'.$$

IV. АКСИОМЫ НЕПРЕРЫВНОСТИ

1. Пусть AB и CD — два произвольных отрезка. Тогда на прямой AB найдется n различных точек $A_1, A_2 \ldots, A_{n-1}, A_n$ таких, что каждый из отрезков $AA_1, A_1A_2, A_2A_3, \ldots, A_{n-1}A_n$ равен отрезку CD, причем точка B лежит между A и A_n.

2. Существует прямая a, обладающая следующим свойством: если $A_1B_1, A_2B_2, A_3B_3, \ldots$ — такая последовательность отрезков на прямой a, что каждый последующий отрезок, начиная со второго, содержится в предыдущем, то на прямой a найдется точка, общая для всех отрезков.

V. АКСИОМА ПАРАЛЛЕЛЬНОСТИ

Пусть a — произвольная прямая, и A — точка, не лежащая на прямой a. Тогда на плоскости, определенной прямой a и точкой A, существует самое большее одна прямая, проходящая через точку A и не пересекающая прямую a.

Геометрическая система, построенная на перечисленных аксиомах, называется *евклидовой геометрией*, так как она совпадает с геометрией, изложенной Евклидом в его «Началах».

ГЕОМЕТРИЯ И ДЕЙСТВИТЕЛЬНОСТЬ

Все факты, которые закреплены в приведенных выше аксиомах Гильберта, а также вытекающие из них геометрические теоремы и следствия хорошо согласуются с опытом и очень точно отражают соотношения, имеющие место в материальном мире, вследствие чего находят весьма широкое практическое применение. Конечно, не может быть и речи о том, чтобы все эти утверждения описывали пространтсвенные соотношения с исчерпывающей полнотой — никакая наука не дает законченного описания предмета своего изучения. К связи геометрии с действительностью мы вернемся несколько позже.

Ту же самую геометрическую систему можно построить и на основе многих других систем аксиом. Известно немало таких примеров. Причина заключается в том, что выбор аксиом до некоторой степени произволен. Ни о каком геометрическом утверждении нельзя сказать с полной определенностью, является ли оно аксиомой или нет. Это зависит лишь от того, принимается ли данное утверждение в конкретном геометрическом построении без доказательства, как основной факт, или же доказывается как теорема.

В процессе развития геометрии немало исследований было посвящено замене изложенных аксиом другими, а также геометриям, построенным на других системах аксиом. С точки зрения математики мы поступили бы безупречно и в том случае, если бы при логическом построении геометрии исходили из такой аксиоматики, которая не описывает окружающего нас пространства. О роли таких геометрических систем в самой математике речь пойдет ниже.

Геометрические системы, отличные от евклидовой, называют *неевклидовыми геометриями.*

ЭЛЕМЕНТАРНАЯ ЕВКЛИДОВА ГЕОМЕТРИЯ

В дальнейшем мы будем рассматривать элементы евклидовой геометрии не аксиоматически, а исходя из наглядных соображений.

ЛУЧ, ОТРЕЗОК, ПОЛУПЛОСКОСТЬ, ПОЛУПРОСТРАНСТВО

Всякая прямая делится любой своей точкой на два *луча* (две *полупрямых*), при этом точка деления называется началом лучей.

Часть прямой, лежащая между двумя ее точками, называется *отрезком.*

Всякая плоскость делится любой прямой, лежащей в этой плоскости, на две *полуплоскости*, сама прямая называется при этом *границей* полуплоскостей. Никакой отрезок, соединяющий две точки полуплоскости, не имеет общих точек с ее границей.

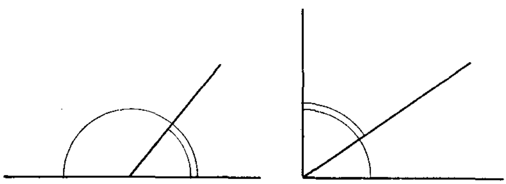
Всякая плоскость делит пространство на два *полупространства.* Две точки тогда и только тогда лежат в одном полупространстве, когда соединяющий их отрезок не имеет общих точек с плоскостью, ограничивающей это полупространство.

ВЗАИМНОЕ РАСПОЛОЖЕНИЕ ПРЯМЫХ И ПЛОСКОСТЕЙ

Две прямые на плоскости *пересекаются*, если они имеют общую точку, и *параллельны*, если не существует точки их пересечения. Если две прямые не лежат в одной плоскости, то они называются *скрещивающимися*. Две плоскости *пересекаются*, если они имеют общую прямую, если же такой прямой нет, то плоскости *параллельны*. Прямая может принадлежать плоскости, пересекать ее в некоторой точке или же не иметь с ней общих точек; в последнем случае говорят, что прямая параллельна плоскости.

УГЛЫ; ПАРЫ УГЛОВ

Два луча, выходящие из некоторой точки плоскости, делят эту плоскость на две части, которые называются *углами*. Лучи, образующие угол, называются *сторонами*, а часть плоскости, соответствующая углу, — *внутренней областью* угла. Внутренняя область угла обозначается буквой или дугой окружности.

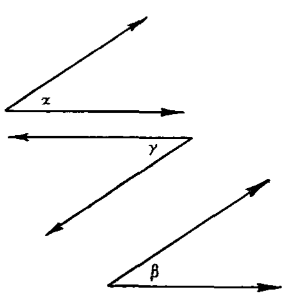


Смежные и дополнительные углы

Два угла называются смежными, если у них одна сторона общая, а две другие составляют прямую линию. Угол, равный смежному с ним углу, называется прямым.

Развернутым называется угол, стороны которого составляют прямую линию. Развернутый угол равен сумме двух прямых углов.

Если два угла в сумме дают прямой угол, то они называются дополнительными друг к другу. Углы, меньшие прямого, называются острыми, а углы, большие прямого, но меньшие развернутого — тупыми; углы, большие развернутого, называют также выпуклыми. Полным называется угол, стороны которого совпадают. Полному углу соответствует одна область — вся плоскость.

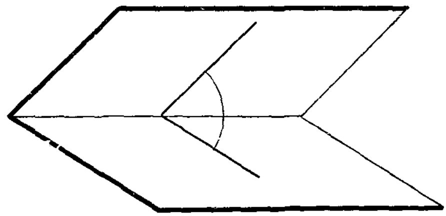


α и β - соответственные углы; пары углов α и γ, β и γ являются накрест лежащими

Если на плоскости вращать луч вокруг его начальной точки, то мы будем получать *углы вращения*. Угол вращения может оказаться и больше полного угла (если, совершив полный оборот, луч продолжает вращение). Угол вращения считается *положительным*, если вращение противоположно направлению движения часовой стрелки, и *отрицательным* — в противном случае.

Два выпуклых угла называются соответственными, если их стороны имеют попарно одинаковые направления, и накрест лежащими, если их стороны имеют попарно противоположные направления. Если накрест лежащие углы имеют общую вершину, то они называются вертикальными.

Соответственные углы так же, как и накрест лежащие, равны между собой.



Угол между двумя пересекающимися плоскостями

Две пересекающиеся прямые образуют две пары равных (вертикальных) углов. Если прямые не являются взаимно перпендикулярными, то за угол между прямыми принимается меньший из этих углов. Под углом между двумя скрещивающимися прямыми понимается угол между прямыми, проведенными параллельно им через некоторую точку пространства. За угол между двумя пересекающимися плоскостями принимается угол между прямыми, проведенными в этих плоскостях перпендикулярно линии их пересечения. Углом между прямой и плоскостью называется угол, который составляет прямая со своей ортогональной проекцией на эту плоскость (если только прямая не перпендикулярна плоскости). Прямая называется перпендикулярной плоскости, если она перпендикулярна любой прямой, лежащей в этой плоскости.

ИЗМЕРЕНИЕ ОТРЕЗКОВ И УГЛОВ

Всякому отрезку можно поставить в соответствие некоторое положительное число (меру) таким образом, чтобы равные отрезки имели одинаковую меру, а сумма мер двух отрезков была равна мере суммарного отрезка, полученного соединением двух исходных отрезков. Мера зависит от выбора отрезка, принимаемого за единицу измерения. Меру отрезка называют также *расстоянием* между двумя его крайними точками. Обратно: после выбора единичного отрезка для любого положительного числа найдется отрезок, мера которого равна этому числу.

За единицу измерения углов принимается одна 90-я доля прямого угла, называемая градусом (1°); 60-я доля градуса называется минутой ($1'$), а 60-я доля минуты — секундой ($1''$). В некоторых местах (особенно на территориях, где основным языком является немецкий) принято использовать так называемый *новый градус* ($1_\circ$), равный сотой доле прямого угла; сотая доля нового градуса называется *новой минутой* ($1_,$), а сотая доля новой минуты — *новой секундой* ($1_{,,}$).

В высшей математике — в первую очередь в теории функций — используется так называемая *радианная мера* углов.

В этом случае мера развернутого угла равна числу π (приближенно 3,1416). Числовое значение радианной меры угла равно длине дуги, высекаемой сторонами угла из единичной окружности с центром в его вершине.

Укажем радианную меру некоторых из наиболее часто встречающихся углов : $360° = 2\pi$, $180° = \pi$, $90° = \frac{\pi}{2}$, $60° = \frac{\pi}{3}$, $45° = \frac{\pi}{4}$. Угол, радианная мера которого равна единице, называется *радианом*.

$$\boxed{1 \text{ радиан} = 57°17'44{,}6''.}$$

Связь между радианной и градусной мерами дается следующей формулой :

$$\boxed{\alpha° = \alpha \cdot \frac{\pi}{180} \text{ радиан}.}$$

МНОГОУГОЛЬНИК, ОКРУЖНОСТЬ; ПОНЯТИЕ ВЫПУКЛОЙ ФИГУРЫ

Многоугольник. Совокупность последовательно соединенных отрезков A_1A_2, A_2A_3, ..., $A_{n-1}A_n$, A_nA_1 на плоскости называется замкнутой ломаной линией, если кроме точек соединения эти отрезки не имеют других общих точек.

Часть плоскости, ограниченная замкнутой ломаной линией, называется *многоугольником*. Отрезки ломаной называются сторонами многоугольника, а точка A_1, A_2, ..., A_n — его вершинами. (Иногда рассматривают также самопересекающиеся ломаные линии. Если не оговорено обратное, то мы будем заниматься лишь так называемыми *простыми многоугольниками*, определенными выше). Если ломаная имеет n вершин, то говорят об n-угольнике (треугольнике, четырехугольнике и так далее).

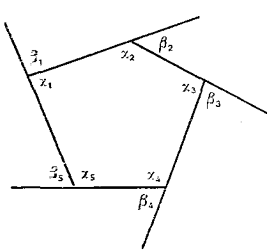
Сумма (внутренних) углов n-угольника равна

$$\boxed{(n-2)\,180°.}$$

Таким образом, для треугольника эта сумма равна 180°, для четырехугольника 360° и так далее. Углы, смежные с углами многоугольника — при условии, что углы многоугольника не превышают 180° — называются внешними углами многоугольника. Сумма внешних углов простого многоугольника равна 360°.

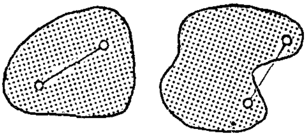
Многоугольник с равными сторонами и углами называется правильным. Внутренний угол правильного n-угольника равен

$$\boxed{\frac{1}{n} \cdot (n-2) \cdot 180°.}$$



Выпуклый многоугольник; углы $\alpha_1, \alpha_2, \ldots$ — внутренние, $\beta_1, \beta_2, \ldots$ — внешние

Окружность. Окружностью называется совокупность всех точек плоскости, лежащих на данном расстоянии от некоторой фиксированной точки той же плоскости. Фиксированная точка называется *центром*, а данное расстояние — *радиусом* окружности. Отрезок, соединяющий две точки окружности, называется *хордой*. Хорда максимальной длины называется *диаметром*, она равна двум радиусам.


Выпуклый и вогнутый элементы плоскости

Выпуклые и вогнутые фигуры. Среди геометрических фигур особое место занимают *выпуклые фигуры*.

Геометрическая фигура называется выпуклой, если вместе с любыми двумя своими точками она содержит и соединяющий их отрезок прямой. Фигуры, не являющиеся выпуклыми, называются вогнутыми.

Выпуклыми фигурами являются, например, прямая, плоскость, луч, полуплоскость, круг, треугольник, угол, не превышающий 180°, шар, куб и другие.

Многоугольник является выпуклым тогда и только тогда, когда все его (внутренние) углы выпуклы (не превышают 180°). Всякий треугольник является выпуклым ; среди четырехугольников уже существуют вогнутые.

ГЕОМЕТРИЧЕСКИЕ ПРЕОБРАЗОВАНИЯ

Одной из важнейших задач геометрических исследований является установление связи между различными геометрическими фигурами, например, выяснение вопроса, каким образом одна фигура может быть получена из другой, то есть какие преобразования нужно произвести с данной фигурой, чтобы получить некоторую другую фигуру. Если производится какое-либо преобразование некоторой фигуры, то интересно исследовать, какие из свойств фигуры при этом изменяются, а какие остаются прежними. Всеми этими вопросами занимается теория *геометрических преобразований*.

О преобразовании мы говорим, когда каждой точке некоторого множества (фигуры) ставится в соответствие некоторая (другая) точка. Точки, которые ставятся в соотвествие точкам исходного множества, называются образами точек исходного множества при данном преобразовании. Назовем преобразование обыкновенным, если различным точкам соответствуют разные образы. Преобразование, не являющееся обыкновенным, называется вырожденным. Преобразования называют также отображениями.

Обыкновенным преобразованием некоторой фигуры является, например, параллельный перенос. Здесь устанавливается соответствие между исходными и сдвинутыми положениями точек. Другой пример : каждой точке кинопленки соответствует одна из точек на экране. Можно установить соответствие между правым и левым ботинками одной пары обуви, каждой точке одного ботинка можно сопоставить соответствующую точку другого. А солнечные лучи, например, осуществляют вырожденное отображение цветка на его тень. Действительно, тень от многих точек цветка падает в одну точку, поэтому соответствие уже не является взаимно однозначным.

В алгебре теория преобразований тесно связана с теорией групп. Классификация преобразований осуществляется согласно тому, какие из свойств фигур они оставляют неизменными, то есть какие свойства являются *инвариатными*. Вследствие этого различают, например, преобразования, сохраняющие расстояния, углы, переводящие окружность в окружность, прямую — в прямую, преобразования, сохраняющие свойство непрерывности и другие.

ПРЕОБРАЗОВАНИЯ СОВМЕЩЕНИЯ (СОХРАНЯЮЩИЕ РАССТОЯНИЕ)

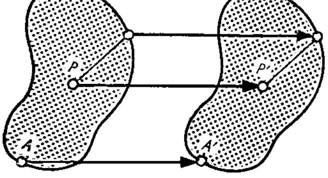
Говорят, что преобразование сохраняет расстояние, если всякому отрезку прямой оно ставит в соответствие равный ему отрезок. Мы назовем такие преобразования совмещениями (преобразованиями совмещения). Две фигуры могут быть совмещены друг с другом, если существует такое преобразование совмещения, которое переводит одну из них в другую.

Совместимые фигуры на плоскости всегда могут быть совмещены друг с другом в результате некоторого движения (для этого, может быть, нужно выйти за пределы плоскости); совместимые же тела — уже не всегда (например, правый и левый ботинки). Совместимость обозначается знаком $\cong$ или $\equiv$. Ниже будут рассмотрены некоторые специальные преобразования совмещения на плоскости.

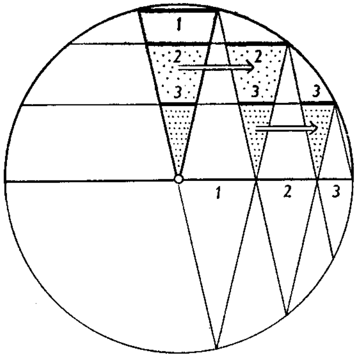
ПАРАЛЛЕЛЬНЫЙ ПЕРЕНОС

Параллельный перенос в некотором смысле является наиболее простым из преобразований совмещения.

При параллельном переносе прямые, соединяющие исходные точки с их образами, параллельны между собой (они задают направление переноса);



Перенос



Сумма отрезков хорд, расположенных между двумя радиусами, равна радиусу окружности расстояния от исходных точек до их ообразов равны (длина переноса).Два последовательно выполняемые переноса могут быть заменены одним, который называется их суммой (или произведением).

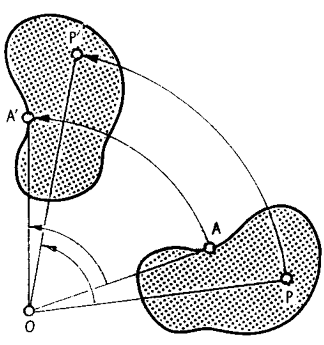
Разделим окружность на нечетное количество равных частей (на рисунке их семь). Точки деления, расположенные на одинаковом расстоянии от диаметра, соединим хордами и проведем радиусы, соответствующие двум средним точкам деления. При помощи переноса нетрудно убедиться, что отрезки хорд, расположенные между этими радиусами, в сумме дают радиус окружности (из рисунка видно, что каждый из этих отрезков можно с помощью максимум двух переносов (вместе с соответствующим треугольником) наложить на радиус).

ПОВОРОТ

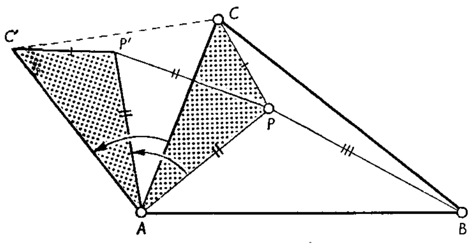
Поворот (вокруг точки) — это такое преобразование совмещения, при котором неподвижной остается только одна точка — центр вращения.

Поворот однозначно определяется заданием центра вращения, направления и угла, на который совершается поворот.

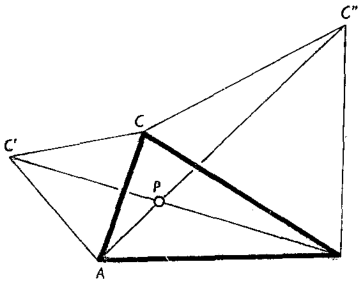
С помощью поворота можно решить, например, следующую задачу: на плоскости требуется найти точку P, сумма расстояний от которой до трех данных точек A,B и C той же плоскости была бы наименьшей. Выберем сначала точку P произвольным образом. Величину суммы $PA + PB + PC$ можно представить наглядно, если отложить эти отрезки последовательно — один от конца другого. Этого можно добиться, если, например, повернуть треугольник APC вокруг точки A «во внешнюю сторону» на 60° (см. рисунок). При этом получатся равносторонние треугольники ACC', и APP', (оба треугольника равнобедренные с углом при вершине, равным 60°). Сумма отрезков $C',P', + P',P + PB$ равна сумме $PA + PB + PC$, причем она достигает максимума, когда все три отрезка лежат на одной прямой. Отсюда следует, что искомая точка P должна лежать на прямой, соединяющей точку B с точкой C', которая является вершиной равностороннего треугольника, построенного на отрезке AC.



Поворот



Определение точки P, которой соответствует наименьшая сумма расстояний до трех данных точек



Построение изогональной точки

Аналогичные соображения позволяют заключить, что точка P должна лежать на прямой, соединяющей точку A с точкой C'', которая является вершиной равностороннего треугольника, построенного на отрезке BC. Следовательно, искомая точка P является точкой пересечения прямых AC'' и BC'. (Предполагается, что ни один из углов треугольника ABC не превышает 120°). Точка P называется *изогональной* точкой треугольника, так как из нее все стороны треугольника видны под одним и тем же углом.

ЗЕРКАЛЬНОЕ ОТРАЖЕНИЕ ОТНОСИТЕЛЬНО ТОЧКИ

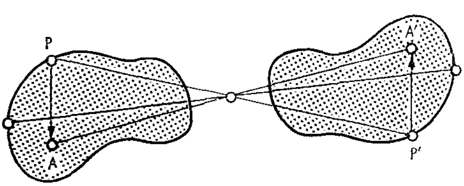
Зеркальное отражение от некоторой точки (симметрия относительно точки) равносильна повороту вокруг этой точки на 180°.

Это преобразование переводит всякую прямую в параллельную ей прямую.

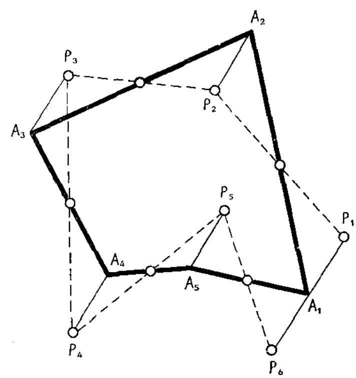
Фигура называется центрально-симметричной, если существует зеркальное отражение относительно некоторой точки, отображающее эту фигуру на себя.

Таковы, например, параллелограмм, окружность, плоскость и так далее.

Рассмотрим на плоскости некоторый многоугольник $A_1A_2 \ldots A_n$ с нечетным числом сторон (на рисунке изображен пятиугольник). Если произвольную точку P_1 той же плоскости соединить с точкой A_1, а затем зеркально отразить отрезок P_1A_1 по очереди от середин каждой из сторон многоугольника, то зеркальные образы P_2A_2, P_3A_3, . . ., P_nA_n будут равны и параллельны отрезку P_1A_1, поэтому последний образ, то есть отрезок



Центральная симметрия



Построение многоугольника с нечетным числом сторон по серединам его сторон

$P_{n+1}A_1$, лежит на одной прямой с отрезком P_1A_1, причем точки P_{n+1} и P_1 расположены симметрично относительно точки A_1. Это дает возможность построить многоугольник с нечетным числом сторон, если заданы лишь середины его сторон. Действительно, если зеркально отразить точку P_1 по очереди от середин всех сторон, то середина отрезка, соединяющего точку P_1 с последним образом, окажется как раз вершиной A_1 искомого многоугольника.

ЗЕРКАЛЬНОЕ ОТРАЖЕНИЕ ОТ ОСИ

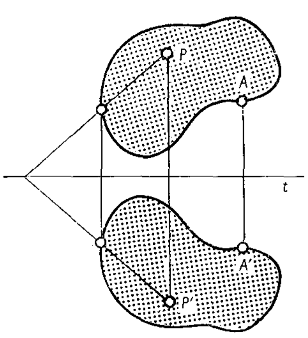
Зеркальное отражение от оси является таким преобразованием совмещения на плоскости, которое всякой фигуре ставит в соответствие ее зеркальный образ относительно этой оси.

Точки, между которыми устанавливается соответствие, лежат по разные стороны от оси на одинаковом расстоянии от нее; соединяющий их отрезок перпендикулярен оси.

Прямая и ее зеркальный образ относительно оси пересекают ось в одной точке или же параллельны ей.

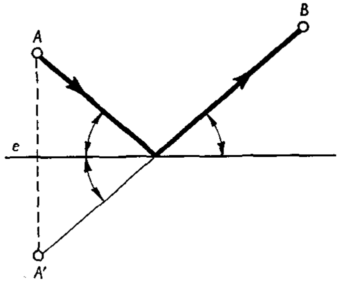
Если существует такое зеркальное отражение относительно некоторой оси, которое отображает фигуру на себя, то эта фигура называется симметричной. Таковы, например, прямая, окружность, квадрат и так далее. В природе симметричные или по крайней мере приближенно симметричные фигуры встречаются очень часто. Примером могут служить цветы, листья и строение организмов живых существ. Такая симметрия часто используется в строительстве, в изобразительном искусстве, при изготовлении украшений и так далее. В геометрии также нередко применяются рассуждения, в которых симметрия играет весьма важную роль.

На плоскости осью симметрии отрезка прямой является перпендикулярная прямая, делящая отрезок пополам, а осью симметрии угла — его *биссектриса*. Таким образом, совокупность точек, каждая из которых одинаково удалена от концов данного отрезка, есть перпендикулярная прямая, делящая отрезок пополам, а совокупность точек, каждая из которых одинаково удалена от сторон данного угла, есть биссектриса угла. Отметим, что совокупность или множество точек, обладающих некоторым общим свойством, часто называют геометрическим местом точек, обладающих данным свойством. Так, например, ось симметрии отрезка прямой на плоскости является геометрическим местом точек, каждая из которых одинаково удалена от концов отрезка.



Осевая симметрия

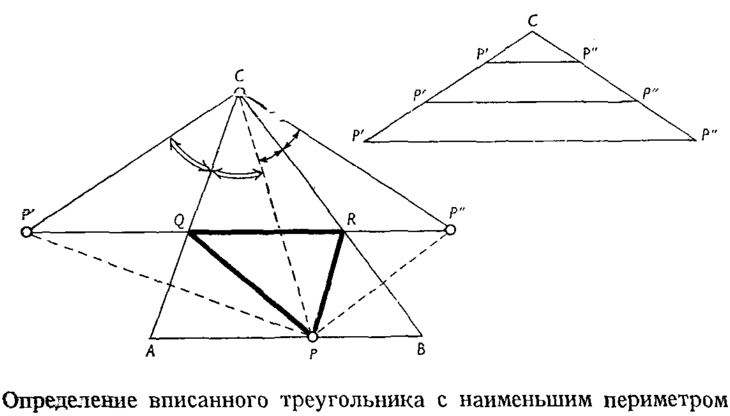
Рассмотрим на плоскости прямую e и точки A и B, лежащие по одну сторону от этой прямой. Определим кратчайший путь, ведущий от A до прямой, а затем в B. Для этого найдем кратчайший путь из точки A' (зеркального образа A относительно прямой e) в точку B. Очевидно, что это отрезок прямой, соединяющий A' и B. Наши рассуждения основаны



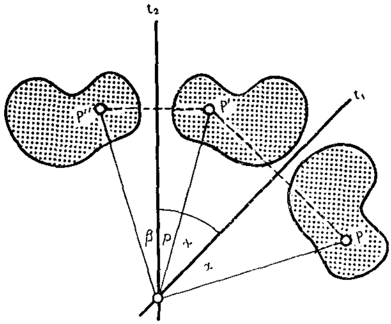
Кратчайший путь из A и B, касающийся прямой e

на том, что расстояние от точки A до любой точки прямой e равно расстоянию от ее зеркального образа A' до той же точки. Нетрудно заметить, что построенный таким образом кратчайший путь совпадает с траекторией светового луча, который, выходя из точки A, отражается от зеркала, проходящего через прямую e перпендикулярно плоскости, а затем попадает в B (угол падения равен углу отражения).

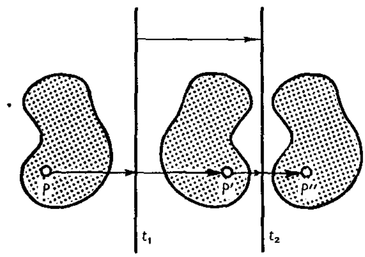
Рассмотрим теперь остроугольный треугольник ABC. Пусть требуется определить кратчайший путь, ведущий из некоторой точки P стороны AB к стороне AC, отсюда к стороне BC, а затем обратно в точку P. Воспользуемся методом, примененным выше. Если P' — зеркальное отражение точки P относительно стороны AC, а точка P'' — относительно стороны BC, то искомый путь равен отрезку $P'P''$. Этот отрезок пересекает стороны треугольника в точках Q и R, следовательно, кратчайшим путем является контур треугольника PQR. Тем самым мы построили треугольник наименьшего периметра, вписанный в данный треугольник (то есть его вершины лежат на сторонах треугольника ABC) и удовлетворяющий тому условию, что одна из его вершин совпадает с данной точкой P. Периметр вписанного таким образом треугольника зависит, естественно, от выбора точки P. Если мы теперь хотим определить положение точки P, при котором треугольник PQR имеет наименьший периметр, то следует принять во внимание, что стороны CP' и CP'' треугольника $P'CP''$ являются зеркальными образами отрезка CP, то есть равны между собой. Отсюда следует также, что $\sphericalangle\ P'CP''$ вдвое больше, чем $\sphericalangle\ ACB$ треугольника ABC, то есть не зависит от выбора точки P. Следовательно, наименьший периметр имеет тот из треугольников $P'CP''$, который имеет наименьшее ребро, то есть для которого соответствующий отрезок CP



Определение вписанного треугольника с наименьшим периметром



Последовательное отражение от пересекающихся осей равносильно повороту



Последовательное отражение от параллельных осей равносильно переносу

имеет наименьшую длину. Таким образом искомая точка P является основанием перпендикуляра, опущенного из точки C на AB.

Среди преобразований совмещения зеркальное отражение от оси играет особую роль. Любое преобразование совмещения может быть получено в результате последовательного выполнения нескольких зеркальных отражений (то есть как *произведение* нескольких зеркальных отражений). Последовательное отражение от двух *пересекающихся прямых* равносильно повороту вокруг точки их пересечения на угол, вдвое превышающий угол между этими прямыми.

Произведение отражений от двух *параллельных осей* равносильно переносу в направлении, перпендикулярном осям, на расстояние, вдвое превышающее расстояние между ними. Две совместимые фигуры на плоскости всегда могут быть переведены одна в другую при помощи самое большее трех последовательных зеркальных отражений.

ПОДОБИЕ, ЦЕНТРАЛЬНОЕ ПОДОБИЕ

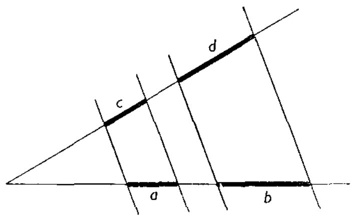
В отличие от преобразований, сохраняющих расстояния, которые изменяют лишь положение фигур (тел) в пространстве, преобразования подобия вызывают большие изменения.

Преобразованием подобия мы называем такое преобразование, при котором отношение образа любого отрезка к самому отрезку постоянно. Это отношение называется коэффициентом подобия. Если коэффициент подобия больше 1, то говорят о подобном расширении, если меньше 1 — о сжатии. Преобразования, сохраняющие расстояния, являются частным случаем преобразований подобия, для них коэффициент подобия равен 1. Всякому углу преобразование подобия ставит в соответствие равный ему угол, то есть *сохраняет углы*.

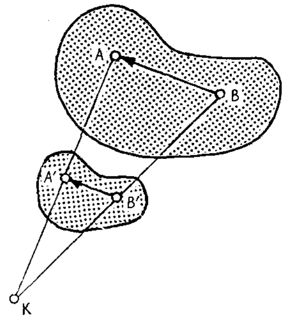
Две фигуры (тела) называются подобными, если существует преобразование подобия, переводящее одну из них в другую.

Подобие обозначается знаком $\sim$.

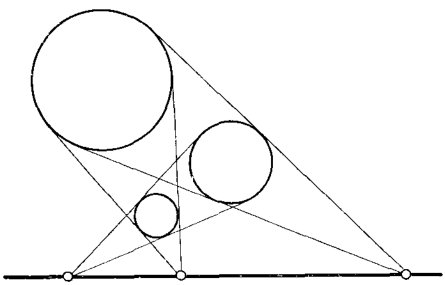
В исследовании подобных фигур основную роль играет *теорема о па-*



Георема о параллельных секущих: $a : b = c : d$



Центральное подобие



Внешние центры подобия трех окружностей лежат на одной прямой

раллельных секущих: если стороны угла пересечь параллельными прямыми, то отношение отрезков, отсеченных на одной из сторон, равно отношению соответствующих отрезков на другой стороне угла.

Если прямые, соединяющие точки с их образами при подобном преобразовании, пересекаются в одной точке, то говорят о центрально-подобном преобразовании (называемом также гомотетией).

Центрально-подобное преобразование переводит всякую прямую в параллельную ей прямую или в саму себя. Последовательное выполнение двух центрально-подобных преобразований равносильно одному центрально-подобному преобразованию или параллельному переносу. Центры подобия трех фигур, каждые две из которых центрально-подобны, лежат на одной прямой.

Две окружности с разными радиусами имеют два центра подобия (внутренний и внешний). Таким образом внешние центры подобия, соответствующие каждым двум из трех данных окружностей, лежат на одной прямой.

РАВЕНСТВО И ПОДОБИЕ ТРЕУГОЛЬНИКОВ

Для того, чтобы два треугольника были равны или подобны, достаточно, чтобы их стороны или углы удовлетворяли определенным условиям. Соответственно случаям, наиболее часто встречающимся на практике, выделяют *основные признаки равенства и подобия треугольников.*

Два треугольника равны, если

а) три стороны одного треугольника соответственно равны трем сторонам другого;

б) две стороны и угол между ними одного треугольника соответственно равны двум сторонам и углу между ними другого треугольника;

в) сторона и два угла одного треугольника соответственно равны стороне и двум углам другого треугольника;

г) две стороны и наибольший из противоположных им углов одного треугольника соответственно равны двум сторонам и наибольшему из противоположных им углов другого треугольника.

Два треугольника подобны, если

а) три стороны одного треугольника пропорциональны трем сторонам другого;

б) две стороны одного треугольника пропорциональны двум сторонам другого, и углы между ними равны;

в) два угла одного треугольника равны двум углам другого;

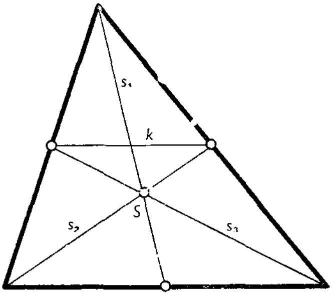
г) две стороны одного треугольника пропорциональны двум сторонам другого, и наибольший из противоположных им углов одного треугольника равен соответственному углу другого.

НЕКОТОРЫЕ НАИБОЛЕЕ ИЗВЕСТНЫЕ ТЕОРЕМЫ О ТРЕУГОЛЬНИКАХ

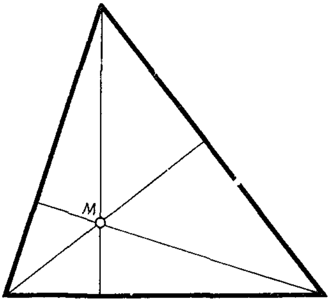
В некотором смысле треугольник является простейшей геометрической фигурой, по этой причине известно множество теорем о его элементах. Некоторые из них имеют и практическое применение.

Отрезок прямой, соединяющий середины двух сторон треугольника, называется его средней линией, он параллелен третьей стороне и равен ее половине.

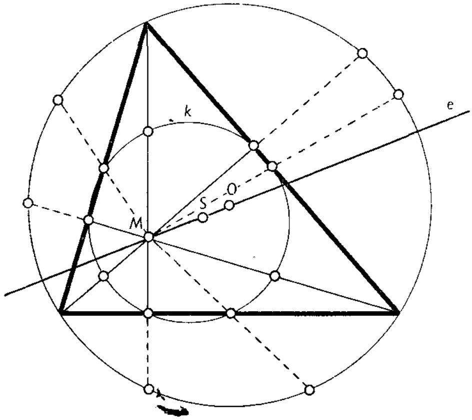
Отрезок прямой, соединяющий середину стороны треугольника с противоположной ей вершиной, называется медианой. Медианы пересекаются в одной точке, называемой центром тяжести треугольника. Центр тяжести отделяет от каждой из медиан ее третью часть.



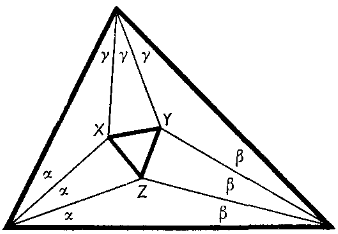
k — средняя линия; s_1, s_2, s_3 — медианы, S — центр тяжести



Точка пересечения высот



Окружность Фейербаха (k) и прямая Эйлера (e)



Теория Морлея: треугольник XYZ — правильный

Перпендикуляр, опущенный из любой вершины треугольника на противоположную сторону, называется высотой. Высоты треугольника пересекаются в одной точке.

Перпендикуляры к сторонам треугольника, проходящие через их середины, пересекаются в центре описанной около треугольника окружности. Биссектрисы внутренних углов пересекаются в центре окружности, вписанной в треугольник.

Если из точки пересечения высот произвести центрально-подобное преобразование описанной около треугольника окружности с коэффициентом подобия $\frac{1}{2}$, то мы получим так называемую окружность Фейербаха, которая обладает следующими свойствами:
она проходит через середины сторон и через основания высот треугольника, а также делит пополам отрезки высот от точки их пересечения до вершин треугольника.

Точка пересечения высот, центр окружности Фейербаха, центр тяжести треугольника и центр описанной около него окружности лежат на одной прямой, называемой прямой Эйлера для данного треугольника.

Теоремы о замечательных точках в треугольнике обобщает *теорема Цева:* если C', A' и B' — точки на сторонах AB, BC и CA треугольника ABC, то отрезки AA', BB' и CC' тогда и только тогда пересекаются в одной точке, когда $\frac{AC'}{C'B}\frac{BA'}{A'C}\frac{CB'}{B'A} = 1$.

Следующая *теорема Морлея* известна скорее благодаря своей математической «красоте»: если в произвольном треугольнике провести линии, делящие каждый из внутренних углов треугольника на три части, то точки пересечения тех из них, которые проходят ближе всего к сторонам треугольника, являются вершинами некоторого равностороннего треугольника.

НЕКОТОРЫЕ НЕРАВЕНСТВА, СВЯЗАННЫЕ С ТРЕУГОЛЬНИКАМИ

Из трех отрезков тогда и только тогда можно построить треугольник, когда сумма любых двух из них больше третьего.

В любом треугольнике против большей стороны лежит больший угол; против равных сторон лежат равные углы. Отсюда следует, что в *равнобедренном треугольнике* есть два равных угла, а в *равностороннем (правильном) треугольнике* равны все три угла (по 60°). В остроугольном треугольнике сумма расстояний от вершин минимальна для изогональной точки, а сумма квадратов расстояний от вершин — для центра тяжести треугольника. Минимальная сумма расстояний от трех сторон треугольника соответствует вершине, противоположной наибольшей стороне. Сумма расстояний от произвольной внутренней точки треугольника до его вершин по крайней мере вдвое превышает сумму расстояний от этой точки до сторон треугольника *(теорема Эрдёша—Морделла)*.

ЧЕТЫРЕХУГОЛЬНИКИ: ТРАПЕЦИИ, ПАРАЛЛЕЛОГРАММЫ

На практике из четырехугольников применяют в первую очередь трапеции и параллелограммы.

Четырехугольник называется трапецией, если две его стороны параллельны, и параллелограммом, если он имеет две пары параллельных сторон.

Таким образом, параллелограмм является частным видом трапеции.

Отрезок, соединяющий середины непараллельных сторон трапеции, называется ее средней линией; средняя линия равна среднему арифметическому параллельных сторон трапеции.

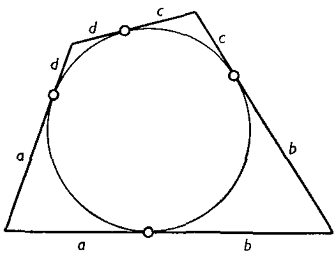
Всякий параллелограмм симметричен относительно точки пересечения своих диагоналей, таким образом его диагонали в точке пересечения делятся пополам. Противоположные углы параллелограмма так же, как и противоположные стороны, равны между собой. Сумма углов, прилежащих к одной стороне параллелограмма, равна 180°.

Параллелограмм, все стороны которого равны между собой, называется ромбом.

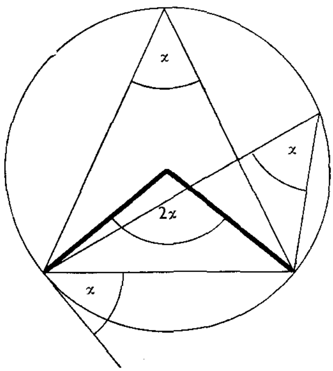
Диагонали ромба взаимно перпендикулярны.

Параллелограмм, все углы которого прямые, называется прямоугольником. Диагонали прямоугольника равны. Фигурой, обладающей и свойствами ромба, и свойствами прямоугольника, является квадрат.

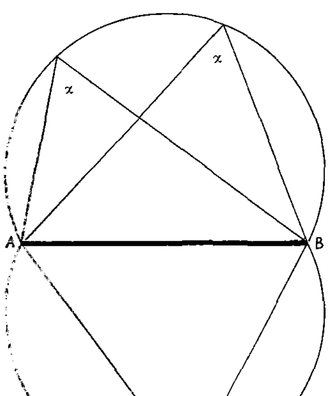
Середины сторон произвольного четырехугольника являются вершинами некоторого параллелограмма. Стороны этого параллелограмма параллельны соответствующим диагоналям четырехугольника. (Раньше параллелограмм, который не является ни ромбом, ни прямоугольником, называли ромбоидом; в наши дни это название уже почти не употребляется).



Описанный четырехугольник



Центральные и вписанные углы



Дуги окружностей, из точек которых данный отрезок виден под одним углом

О ГЕОМЕТРИИ ОКРУЖНОСТИ

Прямая, имеющая с окружностью одну точку, называется касательной к ней в этой точке. Касательная перпендикулярна к радиусу, проведенному в точку касания. Из точки вне окружности можно провести две касательных к ней; их отрезки до точек касания равны между собой. Отсюда следует, что суммых противоположных сторон *описанного около окружности* четырехугольника равны.

Угол между двумя радиусами окружности называется центральным углом.

Если из окружности исключить дугу, соответствующую некоторому центральному углу, то останется дуга, из любой точки которой отброшенная дуга видна под одним и тем же углом. Каждый такой угол называется вписанным углом, соответствующим упомянутому центральному углу.

Вписанный угол всегда равен половине соответствующего центрального угла. Это верно и для угла, одна из сторон которого является хордой, стягивающей соответствующую дугу, а другая — касательной к окружности в одном из концов этой дуги. Действительно, такой угол можно рассматривать как предельное положение вписанного угла. Обратно : рассмотрим на плоскости некоторый отрезок; совокупность точек (лежащих по одну сторону от прямой, на которой лежит отрезок), из которых данный отрезок виден под одним и тем же углом, есть дуга некоторой окружности. Всякая хорда делит окружность на две дуги. Сумма углов, под которыми хорда видна с этих дуг, равна 180°. Отсюда следует, что сумма противоположных углов вписанного четырехугольника равна 180°.

Теорема Талия. Если центральный угол равен 180°, то есть его стороны составляют диаметр окружности, то соответствующие вписанные углы равны 90°. Таким образом *всякий отрезок виден под прямым углом из точек окружности, диаметром которой он является.* Это *теорема Талия.*

ПЛОЩАДЬ

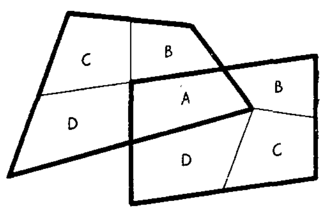
За единицу площади примем площадь единичного квадрата (квадрата с единичной стороной). Площадь многоугольника всегда является положительным числом, совместимые многоугольники имеют одинаковую площадь. Если многоугольник разбить на части, то сумма площадей отдельных частей будет равна площади всего многоугольника. Число, обладающее такими свойтвами, можно поставить в соответствие каждому многоугольнику единственным образом.

Площадь прямоугольника равна произведению его соседних сторон, *площадь параллелограмма* — произведению его стороны на соответствующую высоту, а *площадь треугольника* — половине площади соответствующего параллелограмма (так как каждый треугольник можно рассматривать как половину некоторого параллелограмма). *Площадь треугольника* можно вычислить и по трем его сторонам:

$$\boxed{t = \sqrt{s(s-a)(s-b)(s-c)}}$$

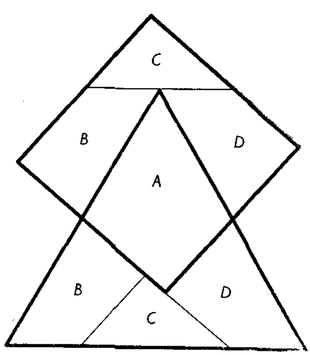
(формула Герона), где a, b и c — стороны треугольника, а s — половина его периметра.

Площадь трапеции равняется произведению ее средней линии на высоту. Если диагонали четырехугольника взаимно перпендикулярны, то его площадь равна половине произведения диагоналей. Для площадей более сложных многоугольников, вообще говоря, нельзя указать таких простых формул. На практике для определения площадей таких многоугольников (например, земельных участков) их разбивают на трапеции и треугольники.



Построение параллелограмма из кусков четырехугольника

Если имеется два равновеликих многоугольника (то есть имеющих одинаковую площадь), то каждый из них можно разбить на конечное число кусков таким образом, что из полученных кусков можно составить другой многоугольник (теорема Фаркаша Бойаи.) Если разрезать произвольный четырехугольник вдоль его средних линий (соединяющих середины противоположных сторон), то из полученных кусков легко составить параллелограмм. Значительно трудней преобразовать таким образом правильный треугольник в квадрат.



Построение квадрата из кусков треугольника

Для определения площади произвольной плоской фигуры (например, фигуры, ограниченной некоторой плоской кривой) введем два понятия.

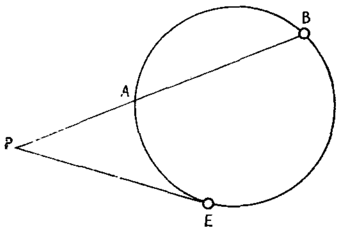
Для данной плоской фигуры назовем внутренним (вписанным) многоугольником всякий многоугольник, целиком содержащийся внутри этой фигуры, а внешним (описанным) многоугольником — всякий многоугольник, содержащий данную фигуру.

Вычисление площадей сводится, по сути дела, к сравнению искомой площади с площадью единичного квадрата, что в случае многоугольников не представляет особых трудностей. Площади фигур более общего вида можно приблизить с помощью площадей многоугольников. Эти соображения приводят к следующему определению: площадью плоской фигуры называется некоторое положительное число, обладающее тем свойством, что для этой фигуры не существует ни внешнего многоугольника с площадью, меньшей, чем это число, ни внутреннего многоугольника с площадью, большей, чем это число. (Площадь определяется лишь для тех фигур, для которых существует только одно такое число). Основываясь на этом определении, можно найти площадь круга: площадь круга радиуса r равна $r^2\pi$. При растяжении плоской фигуры в n раз ее площадь увеличивается в n^2 раз.

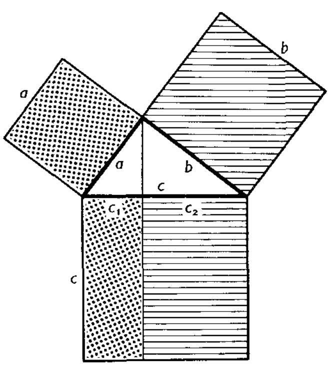
СРЕДНЕЕ ГЕОМЕТРИЧЕСКОЕ, ТЕОРЕМА ПИФАГОРА

Средним геометрическим двух расстояний называется квадратный корень из их произведения. Другими словами, произведение длин двух отрезков равно квадрату их среднего геометрического.

Отрезок касательной к окружности от некоторой точки до точки касания равен среднему геометрическому отрезков, отсекаемых окружностью от секущей, проведенной из той же точки. В прямоугольном треугольнике любой из катетов (сторон прямого угла) равен среднему геометрическому гипотенузы (стороны, противоположной прямому углу) и ортогональной проекции этого катета на гипотенузу. Таким образом, если обозначить катеты через a и b, гипотенузу через c, проекцию a на c через c_1, а проекцию b на c — через c_2, то, например, $a^2 = cc_1$. Это означает, что



Отрезок касательной есть среднее геометрическое отрезков секущей: $PE^2 = PA \cdot PB$



Среднее геометрическое и теорема Пифагора площадь квадрата, построенного на катете a, равна площади прямоугольника со сторонами с и c_1. Аналогично: площадь квадрата, построенного на катете b, равна площади прямоугольника со сторонами c и c_2. Поскольку эти два прямоугольника вместе составляют квадрат, построенный на гипотенузе, то получаем, что для прямоугольного треугольника площадь квадрата, построенного на гипотенузе, равна сумме площадей квадратов, построенных на катетах. В наших обозначениях

$$\boxed{a^2 + b^2 = c^2.}$$

Это соотношение называется *теоремой Пифагора*. Оно позволяет по двум сторонам прямоугольного треугольника определить его третью сторону.

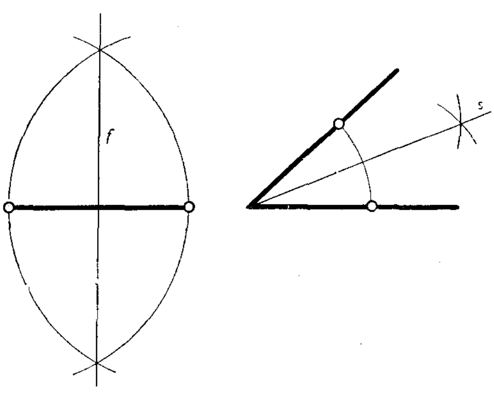
ПОСТРОЕНИЯ

Это одна из наиболее известных областей применения геометрии. Наиболее древними средствами построений являются циркуль и линейка. В процессе развития геометрии были установлены правила применения этих приспособлений при геометрических построениях.

Евклидовыми называются построения, которые могут быть осуществлены в результате выполнения конечного числа основных построений, перечисленных ниже:

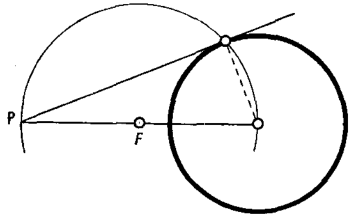
1. проведение прямой через две точки;
2. обозначение точки пересечения двух (пересекающихся) прямых;
3. установление раствора циркуля, равного расстоянию между двумя данными точками;
4. построение окружности данного радиуса с центром в данной точке;
5. обозначение точек пересечения данной прямой с данной окружностью;
6. обозначение точек пересечения двух данных окружностей.

К числу наиболее часто применяемых построений относится построение перпендикуляра, делящего пополам данный отрезок, и построение биссектрисы данного угла. Построение касательной к данной окружности в данной точке можно выполнить, основываясь на теореме Талия, если учесть, что касательная перпендикулярна к радиусу, проведенному в точку касания.



Построение перпендикуляра, делящего отрезок пополам, и биссектрисы угла

Существуют задачи на построение, которые нельзя решить с помощью евклидовых построений. С исторической точки зрения интерес представляет, например,



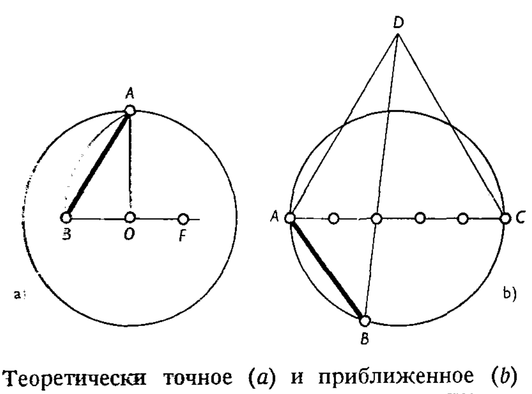
Построение касательной к окружности

задача удвоения куба, то есть построение из ребра данного куба ребра куба, имеющего вдвое больший объем. Другие примеры: в общем случае нельзя построить угол, втрое меньший данного; неразрешима задача квадратуры круга, то есть построение квадрата, площадь которого равнялась бы площади данного круга.

Общий метод (кстати сказать, не такой уж простой), позволяющий установить, разрешима ли данная задача на построение с помощью евклидовых построений, дается в алгебре (теория поля). Однако на практике нет смысла придерживаться евклидовых построений, существуют и широко применяются другие средства и приемы построений. Проводя, например, параллельные прямые с помощью двух треугольников, мы выходим за рамки евклидовых построений, однако это значительно проще и практически точнее, чем проведение параллельных прямых с помощью евклидовых построений.

Применение других средств построений позволяет иногда решить задачи, неразрешимые в евклидовом смысле.

Приближенные построения. Часто бывает достаточно выполнить так называемое *приближенное построение*, которое, хотя и является теоретически неточным, нередко оказывается не хуже, чем теоретически точное построение, осуществленное на практике. В качестве примера укажем сначала теоретически точное, а затем приближенное построение правильного пятиугольника. На рисунке *a* отрезок AB является стороной вписанного в окружность пятиугольника (дуга AB есть дуга окружности с центром в точке F). На рисунке *b* отрезок AB является приближенной стороной вписанного пятиугольника. Во втором случае точка D является вершиной правильного треугольника, построенного на диаметре AC, который разбит на пять равных частей; прямая DB проходит через вторую точку деления, считая от A.



Теоретически точное (*a*) и приближенное (*b*) построение правильного пятиугольника

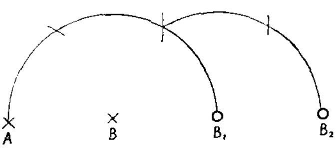
Заметим, что правильный n-угольник можно построить с помощью евклидовых построений только тогда, когда в разложении числа n на простые множители кроме числа 2 присутствуют только числа вида 2^k+1, да и то только в первой степени (Гаусс). Согласно этому, нельзя построить правильные 7—, 9— и 11-угольники, но можно построить правильный 17-угольник. Однако и в случаях, когда евклидово построение неосуществимо, метод, указанный выше для пятиугольника, позволяет выполнить хорошее приближенное построение (в общем случае диаметр AC нужно делить на n равных частей).

ПОСТРОЕНИЯ ОДНИМ ЦИРКУЛЕМ, ОДНОЙ ЛИНЕЙКОЙ

Теоретический интерес представляет собой и тот факт, что циркуль является более действенным приспособлением для построений, чем линейка. Точнее это означает, что — в отличие от линейки — с помощью циркуля можно выполнить любое евклидово построение. Это теорема *Мора—Маскерони*.

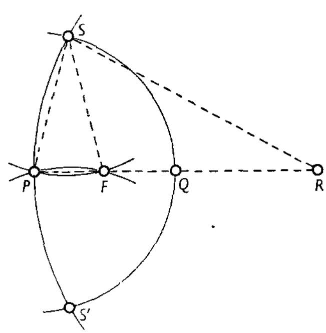
При построениях, выполняемых *одним лишь циркулем*, прямые задаются двумя своими точками. Для доказательства теоремы Мора—Маскерони достаточно показать, что перечисленные выше основные евклидовы построения (кроме первого, которое в этом случае не имеет смысла) выполнимы с помощью одного лишь циркуля.

Не представляет труда построение при помощи одного только циркуля отрезка, кратного данному отрезку AB. Трудней построить точку F, делящую пополам данный отрезок PQ. Построим отрезок PR, вдвое превышающий PQ, и опишем из точки R окружность радиуса PR, а из точки P — окружность радиуса PQ. Искомая точка F будет одной из точек пересечения окружностей радиуса PQ с центрами в точках S и S'. Для доказательства достаточно воспользоваться тем, что равнобедренные треугольники SPR и PFS подобны, и основание каждого из них равно половине боковой стороны.



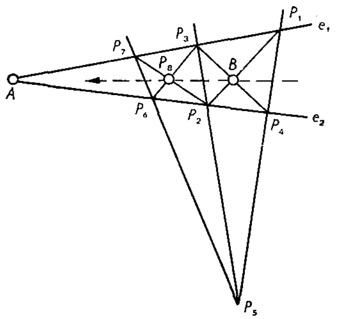
Построение отрезка, кратного данному отрезку, при помощи одного лишь циркуля

С помощью *одной лишь линейки* можно решить относительно мало задач на построение. Однако любое евклидово построение можно вюполнить и с помощью одной лишь линейки, если в плоскости, где выполняется построение, задать некоторую окружность с центром, которую можно использовать в процессе построения. *Штейнер* показал, что если центр окружности не задан, то с помощью одной лишь линейки уже нельзя выполнить любое евклидово построение — не удается, например, найти центр данной окружности.



Деление отрезка пополам при помощи одного лишь циркуля

В рассмотренных построениях мы предполагали, что в нашем распоряжении имеется бесконечно длинная линейка, то есть, что мы можем провести прямую через две точки, удаленные друг от друга на произвольное



Построение прямой, соединяющей точки A и B, с помощью сколь угодно короткой линейки

расстояние. На практике это, конечно, неосуществимо. Оказывается, однако, что можно обойтись и *линейкой конечной длины*, что обеспечивается следующей теоремой: две сколь угодно далекие друг от друга точки можно соединить при помощи сколь угодно короткой линейки.

В самом деле, пусть требуется соединить точки A и B. Любой отрезок можно неограниченно продолжить с помощью сколь угодно короткой линейки, поэтому из точки A в направлении точки B можно провести два луча таким образом, чтобы они проходили в достаточной близости от B и чтобы точка B лежала между ними. Пересечем эти лучи двумя прямыми, проходящими через точку B, в результате чего получим, соответственно, точки P_1, P_2, ... Если теперь из точки P_5 провести прямую P_6P_7, проходящую в достаточной близости от B, то мы получим такую точку P_8, что прямая P_8B проходит через A. Корректность проведенного построения вытекает из теоремы Дезарга (см. стр. 329).

МНОГОГРАННИКИ

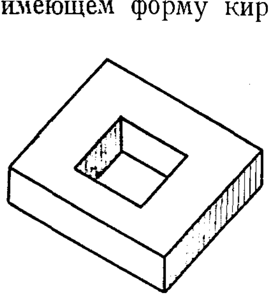
В пространстве многоугольникам соответствуют *многогранники*.

Многогранником мы называем тело, ограниченное конечным числом плоских многоугольников (граней) и не содержащее целиком никакой прямой.

Поверхность многогранника называется связной, если любые две ее точки можно соединить линией, целиком принадлежащей этой поверхности.

Связная поверхность называется односвязной, если любая принадлежащая ей замкнутая кривая делит поверхность на две части.

Неодносвязную поверхность можно получить, если, например, в бруске, имеющем форму кирпича, проделать квадратное отверстие.



Многогранник с неодносвязной поверхностью

Многогранник называется обыкновенным, если всякое его ребро принадлежит двум граням. Нас будут интересовать в первую очередь так называемые простые многогранники, то есть обыкновенные многогранники, поверхности которых односвязны, а грани представляют собой простые многоугольники.

Наглядно это можно сформулировать следующим образом: если простой многогранник изготовить из какого-нибудь эластичного материала, то, «надув» его, можно получить шар, а грани можно путем расширения превратить в круги.

Теорема Эйлера. Всякий выпуклый многогранник является простым. Число граней (l), вершин (c) и ребер (e) простого многогранника удовлетворяют так называемому *соотношению Эйлера:*

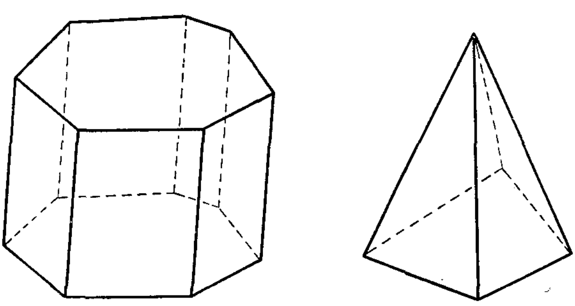
$$\boxed{l + c = e + 2.}$$

Таким образом два числа из этих трех чисел однозначно определяют третье.

К числу наиболее известных многогранников относятся *призмы* и *пирамиды.*

Призма. Если провести через все точки контура некоторого многоугольника прямые, параллельные данному направлению, не параллельному плоскости многоугольника, то мы получим бесконечную призму. Две параллельные плоскости, не параллельные выбранному направлению, отсекают от нее многогранник, называемый призмой. Куски этих плоскостей, отсекаемые призмой, называются основаниями, а расстояние между ними — высотой призмы. Боковые грани призмы представляют собой параллелограммы и в совокупности составляют боковую поверхность призмы. Призма, основания которой также являются параллелограммами, называется параллелепипедом. Если ребра параллелепипеда, выходящие из одной вершины, попарно перпендикулярны, то параллелепипед называется прямоугольным; если эти ребра к тому же равны между собой, то имеем куб.

Пирамида. Проведем из некоторой точки пространства лучи, проходящие через все точки контура некоторого многоугольника. Если выбранная точка не лежит в плоскости многоугольника, то мы получим бесконечную пирамиду. Всякая плоскость, не параллельная ни одному из построенных лучей, отсекает от нее многогранник, называемый пирамидой. Если же пересечь бесконечную пирамиду двумя параллельными плоскостями, не параллельными ни одному из лучей, то получится усеченная пирамида. Куски, отсекаемые при этом от секущих плоскостей, называются основаниями пирамид. Высотой пирамиды называется расстояние от ее вершины (точки, из которой исходят лучи)



Призма и пирамида

до основания; высотой усеченной пирамиды — расстояние между ее основаниями.

Отрезки прямых, использованных при определении призмы и пирамиды, лежащие на поверхности этих многогранников, называются образующими.

ТЕТРАЭДР

Простейшим многогранником является тетраэдр (пирамида, основанием которой служит треугольник). Тетраэдр ограничен четырьмя треугольными гранями. В определенном смысле тетраэдр в пространстве — аналогично треугольнику на плоскости — играет роль элементарного тела; таким образом тетраэдр можно рассматривать как обобщение треугольника. Действительно, многие теоремы о треугольниках могут быть перенесены на тетраэдры.

Плоскости, проходящие через середины ребер тетраэдра и перпендикулряные к ним, пересекаются в одной точке, которая является центром шара, описанного около тетраэдра. Плоскости, делящие пополам внутренние двугранные углы тетраэдра, пересекаются в центре вписанного в него шара. Отрезки, соединяющие вершины тетраэдра с центрами тяжести противоположных граней, пересекаются в центре тяжести тетраэдра; при этом от каждого из отрезков отсекается его четвертая часть. Высотами тетраэдра называются перпендикуляры, опущенные из вершин на противоположные им грани. Вообще говоря, высоты тетраэдра не пересекаются; точнее, они пересекаются только тогда, когда противоположные ребра тетраэдра взаимно ортогональны; в этом случае тетраэдр называется *ортоцентрическим*.

ПРАВИЛЬНЫЕ МНОГОГРАННИКИ

Правильным называется выпуклый многогранник, имеющий равные ребра, углы между ними, а также одинаковые грани.

Одновременно это означает, что его грани представляют собой *равные правильные многоугольники*. Нетрудно убедиться, что существует всего пять видов правильных многогранников. Действительно, если мы хотим построить часть правильного многогранника, прилежащую к одной из его вершин, то можем использовать для этого только правильные треугольники, квадраты или пятиугольники. Дело в том, что внутренний угол правильного шестиугольника равняется 120°; поскольку для построения части многогранника, прилежащей к одной вершине, требуются по меньшей мере три многогранника, то в этом случае мы получили бы в сумме угол 360°, что невозможно, так как во всяком (выпуклом) многограннике сумма углов между ребрами, выходящими из одной вершины, $s < 360°$. Построить часть правильного многогранника, прилежащую к одной вершине так, чтобы сумма углов между соответствующими ребрами не превысила 360°, можно с помощью 3-х, 4-х или 5-ти треугольников, а также с помощью 3-х четырех- или пятиугольников.

В этих пяти случаях построение действительно выполнимо. Соответственно этому известны правильные многогранники следующих типов:

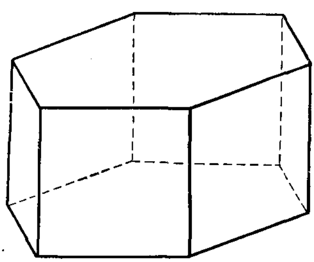
правильный тетраэдр, октаэдр, гексаэдр (или куб), додекаэдр и икосаэдр. Для каждого правильного многогранника можно построить вписанный и описанный шары. Обычно эти многогранники производят впечатление «правильных», благодаря своим свойствам симметрии.

Ниже приведена таблица, содержащая основные данные о правильных многогранниках. Через l, c, e обозначены, соответственно, число граней, вершин и ребер; n обозначает число ребер, принадлежащих одной грани.

	l	c	e	n
правильный тетраэдр	4	4	6	3
октаэдр	8	6	12	3
икосаэдр	20	12	30	3
гексаэдр	6	8	12	4
додекаэдр	12	20	30	5

Как обобщение правильных многогранников рассматривают так называемые полуправильные многогранники. К ним относятся, например, выпуклые многогранники, грани которых представляют собой правильные многоугольники (не обязательно равные), а многогранные углы при вершинах равны.

Такие многогранники могут иметь не более трех граней с различным числом сторон. Простейшим примером полуправильного многогранника является призма, основания которой — правильные n-угольники, а боковые грани — правильные четырехугольники, то есть квадраты.

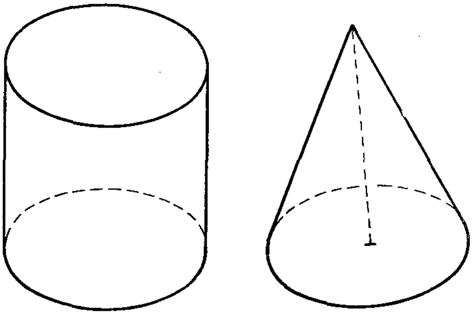


Полуправильное тело

ЦИЛИНДР, КОНУС

Среди тел, ограниченных (кривыми) поверхностями, призмам и пирамидам соответствуют цилиндры и конусы.

Цилиндр. Если через все точки замкнутой плоской кривой (так называемой направляющей) провести прямые, параллельные данному направлению, не параллельному плоскости кривой, то мы получим (бесконечную) цилиндрическую поверхность; тело, которое отсекают от нее две параллельные плоскости, не параллельные выбранному направлению, называют цилиндром. Отрезки прямых, лежащие на поверхности цилиндра, называются его образующими; в совокупности они составляют боко-



Цилиндр и конус вращения

вую поверхность цилиндра. Куски параллельных плоскостей, отсекаемые цилиндрической поверхностью, называются основаниями цилиндра. Если основания — круги, то цилиндр называется круговым; если, кроме того, образующие перпендикулярны основаниям, то имеем прямой круговой цилиндр. Последний называют также цилиндром вращения, так как он может быть получен в результате вращения некоторого прямоугольника вокруг его средней линии.

Пусть r — радиус основания цилиндра вращения, а m — его высота; тогда площадь боковой поверхности цилиндра равна $2r\pi m$. Вычисление площадей боковых поверхностей цилиндров другого типа обычно оказывается значительно более трудной задачей.

Конус. Если провести прямые через каждую точку некоторой плоской кривой (направляющей) и некоторую точку, не лежащую в плоскости направляющей, то мы получим (бесконечную) коническую поверхность; построенные прямые называются образующими конуса (очень часто, а в высшей математике почти всегда, конусом называется полученная поверхность).

В обиходе под конусом понимают тело, которое получается в результате пересечения одной из «половин» бесконечной конической поверхности некоторой плоскостью. Конус вращения получается при вращении одной из двух пересекающихся прямых вокруг другой; пересекая его плоскостью, перпендикулярной оси вращения, можно получить *прямой круговой конус*.

Площадь боковой поверхности прямого кругового конуса равна

$$\boxed{\pi r a,}$$

где a — длина образующей, а r — радиус круга, лежащего в основании конуса.

Тело, отсекаемое двумя параллельными плоскостями от одной из половин конуса вращения, называется усеченным конусом.

Если r, R — радиусы оснований и a — длина образующей, то боковая поверхность усеченного конуса равна

$$\boxed{(R+r)\pi\cdot a.}$$

ГЕОМЕТРИЯ ШАРА

Шаровой или сферической поверхностью (сферой) называется геометрическое место точек, равноудаленных от одной точки, называемой центром шара.

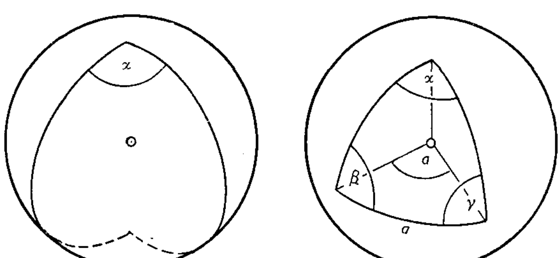
Отрезки, соединяющие центр с точками поверхности, называются радиусами шара. Всякое сечение шара плоскостью есть круг; наибольшие, так называемые *большие круги*, получаются при пересечении шара плоскостями, проходящими через его центр. *Касательная плоскость* имеет с шаром одну общую точку; радиус, проведенный в эту точку, перпендикулярен к касательной плоскости. Если две точки шаровой поверхности не являются противоположными (не лежат на концах одного диаметра), то через них можно провести только один большой круг; через противоположные точки можно провести бесконечно много таких кругов.

Сферический двухугольник. Две полуокужности больших кругов, проходяие через две противоположные точки шаровой поверхности, задают сферический двухугольник; угол между плоскостями этих кругов называют углом сферического двухугольника. (Если не оговорено противное, то берется меньший из двух возможных углов).

Сферический треугольник. Выберем на сфере три точки, которые принадлежат одному большому кругу и никакие две из которых не являются противоположными. Если попарно соединить их кратчайшими дугами окружностей больших кругов, то мы получим сферический треугольник.

Сторонами такого треугольника являются дуги окружностей больших кругов, а его углы равны углам сферических двухугольников, содержащих соответствующие стороны треугольника, то есть равны углу между плоскостями соответствующих больших кругов. Стороны сферического треугольника измеряют соответствующими им центральными углами.

Таким образом, на сфере можно построить так называемую *сферическую геометрию*, напоминающую обычную геометрию на плоскости; роль прямых здесь играют окружности больших кругов. Существенное различие между этими геометриями заключается в том, что на сфере любые две окружности больших кругов (то есть сферические прямые) имеют две общие точки. В то же время и на сфере выполняется, например, то, что дуги, делящие пополам внутренние углы треугольника, пересекаются в центре вписанной окружности, а дуги, перпендикулярные сторонам в их серединах — в центре окружности, описанной около треугольника; медианы также пересекаются в одной точке. (Точнее: вместо центра здесь следует говорить о центрах, так как под центром окружности на сфере следует понимать такую точку сферы, которую можно соединить со всеми точками окруж-



Сферические двухугольник и треугольник

ности дугами одинаковой длины. Однако всякой окружности на сфере соответствуют две такие точки, противоположные друг другу).

Шаровой пояс. Часть шаровой поверхности, заключенная между двумя параллельными секущими плоскостями, называется шаровым поясом; расстояние между этими плоскостями называется высотой пояса. Если одна из двух секущих плоскостей является касательной к шару, то получается сегментная поверхность.

Поверхность шарового пояса, (сегментную поверхность) можно определить, основываясь на теории пределов. Если шар имеет радиус r, то поверхность шарового пояса (сегментная поверхность) высоты m равна

$$\boxed{2r\pi \cdot m.}$$

Поверхность шара. Поверхность шара можно рассматривать как сегментную поверхность, имеющую высоту $2r$. Таким образом поверхность шара равняется

$$\boxed{2r\pi \cdot 2r = 4r^2\pi.}$$

О сферических треугольниках. Рассмотрим теперь на шаре радиуса r сферический треугольник, углы которого, измеренные в радианах, равны α, β и γ. Поверхность этого треугольника определяется по формуле

$$\boxed{F = r^2(\alpha + \beta + \gamma - \pi).}$$

Таким образом (в отличие от геометрии на плоскости) на данном шаре поверхность треугольника однозначно определяется его углами. Поскольку поверхность выражается некоторым положительным числом, то выражение в скобках должно быть положительным, поэтому

$$\boxed{\alpha + \beta + \gamma > \pi.}$$

Другими словами, сумма углов сферического треугольника всегда больше π (или 180°). Отсюда следует также, что сумма углов сферического треугольника не является постоянной. Можно показать, что два сферических треугольника, имеющие соответственно равные углы, равны (конечно, если они лежат на поверхности шаров, имеющих одинаковые радиусы).

ПОНЯТИЕ ОБЪЕМА

Под объемом многогранника понимается положительное число, удовлетворяющее следующим условиям:

1. Объем куба с единичным ребром равен 1.

2. Равные многогранники имеют равные объемы.

3. Объем многогранника равен сумме объемов отдельных его частей.

Основываясь на этих допущениях, можно вывести следующие соотношения для вычисления объемов некоторых тел:

объем *прямоугольного параллелепипеда* равен произведению трех его ребер, выходящих из одной вершины;

объем произвольного *параллелепипеда* равен произведению площади любой его грани на соответствующую высоту;

объем *призмы* равен произведению основания на высоту;

объем *пирамиды* втрое меньше объема призмы с таким же основанием и такой же высотой, то есть равен одной трети произведения площади основания на высоту.

Пусть T и t обозначают, соответственно, площади двух оснований *усеченной пирамиды*, а m — ее высоту. Тогда ее объем равен

$$\boxed{\frac{m}{3}(T + \sqrt{Tt} + t).}$$

Для вычисления объемов других многогранников их обычно разбивают на многогранники перечисленных выше типов, а затем суммируют объемы этих частей.

Мы видели, что каждый из двух равновеликих многоугольников удается разбить на куски, из которых можно составить другой многоугольник. Произвести аналогичное разбиение для равновеликих многогранников, вообще говоря, не удается. Если рассмотреть, например, некоторый куб и выбрать тетраэдр такого же объема, то ни одно из этих тел нельзя разбить на конечное число кусков таким образом, чтобы из них можно было составить другое тело.

Объемы тел, ограниченных (кривыми) поверхностями, определяют, исходя из таких же соображений, как при определении площадей плоских фигур, ограниченных кривыми линиями.

Объем *цилиндра* равен произведению площади его основания на высоту, объем *конуса* — третьей части такого же произведения. Объем *усеченного конуса*, радиусы оснований которого равны R и r, а высота — m, равняется

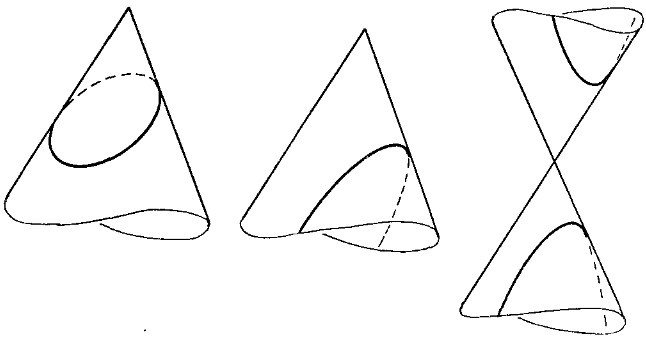
$$\frac{m\pi}{3}(R^2 + Rr + r^2).$$

Объем *шара* радиуса r есть $\dfrac{4r^3\pi}{3}$.

ПЛОСКИЕ СЕЧЕНИЯ КОНУСА

Плоскими сечениями конуса являются все кривые, которые образуются в результате сечения конуса вращения плоскостями.

Эти кривые встречаются и используются в математике и различных приложениях (например, в физике) чаще других и поэтому занимают среди *плоских кривых* особое место.



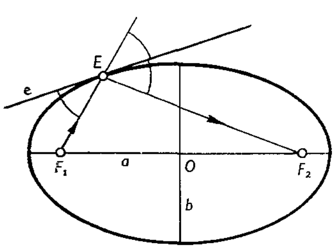
Плоские сечения конуса вращения: эллипс, парабола и гипербола

Если плоскость пересекает все образующие конуса, то в сечении получается эллипс или окружность; если плоскость параллельна одной из образующих, то имеем параболу, если двух — гиперболу.

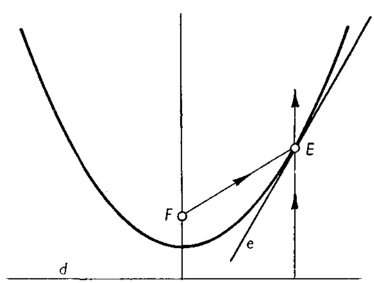
Кривые этих трех типов можно определить и как множества точек, обладающих определенными свойствами.

Эллипс. Эллипс есть геометрическое место точек, для каждой из которых сумма расстояний от двух данных точек (фокусов) есть величина постоянная.

Эллипс имеет две оси симметрии, большую и малую, их длины принято обозначать, соответственно, через $2a$ и $2b$. Фокусы лежат на большой оси. Сумма расстояний от любой точки эллипса до фокуса равна $2a$. Эллипс симметричен и относительно точки пересечения осей. Касательная к



e — касательная к эллипсу; F_1 и F_2 — фокусы



Парабола; d — директриса, F — фокус, e — касательная эллипсу делит пополам внешний угол треугольника, вершинами которого являются фокусы и точка касания. Отсюда вытекает важное физическое следствие: пусть некоторая поверхность (зеркало) имеет в поперечном сечении форму эллипса; тогда волны (например, звуковые или световые), выходящие из одного фокуса, после отражения от поверхности собираются в другом фокусе. Окружность радиуса a с центром в точке пересечения осей эллипса называется *главной окружностью*, соответствующей данному эллипсу.

Парабола. Парабола есть геометрическое место точек, для каждой из которых расстояние от данной точки (фокуса) равно расстоянию до данной прямой (директрисы), не проходящей через фокус.

Парабола имеет одну ось симметрии; фокус лежит на этой оси.

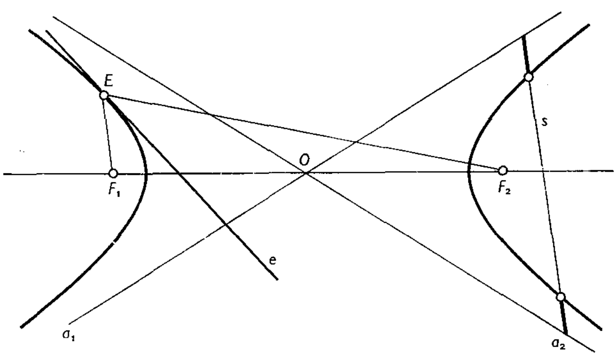
Расстояние от фокуса до директрисы называется параметром параболы и обозначается через p. Вектор, соединяющий фокус с произвольной точкой параболы, называетя радиус-вектором этой точки.

Касательная к параболее делит пополам угол, образованный радиус-вектором точки касания и прямой, проходящей через эту точку параллельно оси параболы. Это имеет важное физическое следствие: если некоторая поверхность (зеркало) имеет в поперечном сечении форму параболы, то волны, выходящие из фокуса, после отражения от поверхности распространяются параллельно оси; обратно: волны, падающие на поверхность параллельно оси, после отражения собираются в фокусе.

Все параболы подобны между собой, то есть существует преобразование подобия, переводящее одну из двух данных парабол в другую.

Гипербола. Гипербола есть геометрическое место точек, для каждой из которых абсолютная величина разности расстояний от двух данных точек (фокусов) есть величина постоянная.

Гипербола имеет две оси симметрии. Одна из них, так называемая *действительная ось*, пересекает гиперболу; другая ось называется *мнимой*. Длину отрезка действительной оси между точками гиперболы принято обозначать через $2a$. Касательная к гиперболе делит пополам внутренний угол треугольника, вершинами которого являются фокусы и точка каса-



Гипербола; F_1 и F_2 — фокусы, a_1 и a_2 — асимптоты, s — секущая, e — касательная ния. Гипербола симметрична и относительно точки пересечения осей. Прямые, проходящие через эту точку, можно разделить на две группы согласно тому, имеют ли они с гиперболой общие точки или нет. Особо выделяют две из этих кривых — так называемые *асимптоты* — которые образуют две пары вертикальных углов, то есть делят плоскость на две области, в одной из которых все упомянутые выше прямые пересекают гиперболу, а в другой — не имеют с ней общих точек. Касательные к гиперболе, параллельные мнимой оси, пересекают асимптоты в точках, расстояние между которыми принято обозначать $2b$.

Отрезки произвольной секущей, лежащие между гиперболой и ее асимптотами, равны между собой. Отрезок касательной к гиперболе, отсекаемый асимптотами, делится в точке касания пополам. Площадь фигуры, ограниченной асимптотами и произвольной касательной, есть величина постоянная. Для данной гиперболы постоянной величиной является и произведение расстояний от произвольной ее точки до асимптот.

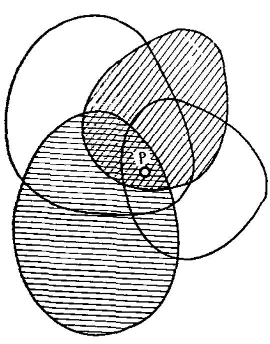
Можно доказать и утверждение, обратное исходному: все рассмотренные кривые могут быть получены в результате сечения конуса вращения некоторыми плоскостями.

ВЫПУКЛЫЕ ОБЛАСТИ НА ПЛОСКОСТИ

В природе многоугольники встречаются сравнительно редко, гораздо чаще встречаются плоские фигуры, ограниченные кривыми. По этой причине целесообразно перейти к исследованию плоских фигур более общего вида.

К таким фигурам относятся выпуклые области на плоскости, обладающие тем свойством, что вместе с любыми двумя своими точками содержат и соединяющий их отрезок.

Сюда же следует отнести и выпуклые многоугольники, однако, в дальнейшем мы не будем делать различия между многоугольниками и областями, ограниченными кривыми.



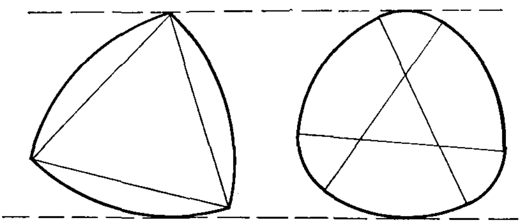
Теорема Хэлли для четырех областей; P — общая точка областей

Пересечение произвольного числа выпуклых областей также есть выпуклая область. Согласно этому, пересечение полуплоскостей, например, является выпуклой фигурой, больше того, всякий выпуклый многоугольник может быть представлен как пересечение полуплоскостей. Характерным свойством выпуклых областей является то, что всякая прямая, проходящая через внутреннюю часть области, пересекает ее границу в двух точках.

Теорема Хэлли. Если любые три из данных n выпуклых областей имеют общую точку, то существует по крайней мере одна точка, общая для всех n областей. Из этой теоремы следует, что если на плоскости имеются n точек, любые три из которых попадают внутрь некоторого круга радиуса

r, то все n точек попадают внутрь некоторого круга радиуса r. В самом деле, выберем произвольным образом три из данных n точек: A, B и C. Центр K круга, покрывающего эти точки, лежит от точек A, B и C на расстоянии, меньшем r, поэтому K будет общей точкой кругов радиуса r с центрами в точках A, B и C. Другими словами, если вокруг каждой из данных n точек построить круги радиуса r, то любые три из этих кругов имеют общую точку. По теореме Хэлли отсюда следует, что существует точка O, являющаяся общей для всех кругов, а это значит, что расстояние от точки O до каждой из исходных точек меньше r. Следовательно, круг радиуса r с центром в точке O покроет все n точек.

Назовем опорной прямой для данной выпуклой области такую прямую, которая имеет с областью по крайней мере одну общую точку, и вся область лежит по одну сторону от этой прямой.


Области постоянной ширины

Для всякой выпуклой области можно построить две опорные прямые, перпендикулярные данному направлению; расстояние между этими прямыми мы будем называть *шириной* области, соответствующей данному направлению. Если всякому направлению соответствует одна и та же ширина, то говорят, что область имеет *постоянную ширину*. Простейшим примером такой области является круг. Область постоянной ширины можно получить также, если около каждой из сторон правильного треугольника описать дугу окружности с центром в противоположной вершине. Такие области находят широкое применение в технике (например, в проекторах). Если область постоянной ширины «катить» по прямой, то при этом ее верхняя точка будет оставаться на постоянном расстоянии от этой прямой. Хотя такие области могут быть построены не только из дуг окружностей, длина границы любой области ширины d равна длине окружности диаметра (то есть ширины) d.

Для выпуклых областей характерно и то, что из любой точки такой области «видна» вся ее граница, то есть отрезки, соединяющие произвольную точку области со всеми точками ее границы, целиком принадлежат области. Исходя из этого свойства, можно прийти к обобщению понятия выпуклой области, а именно: к понятию *звездообразной области*.

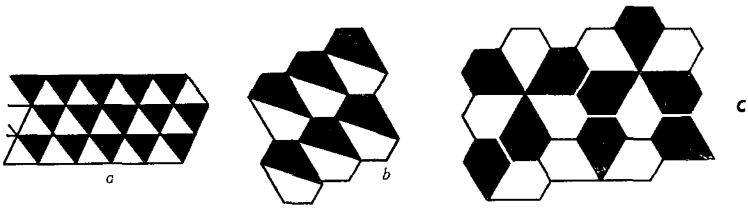

Звездообразная область

Назовем область звездообразной, если у нее существует хотя бы одна внутренняя точка, из которой видны все точки границы области.

Имеет место следующая интересная теорема: если область такова, что для любых трех точек ее границы найдется такая внутренняя точка, из которой видны все три точки, то область является звездообразной.

МОЗАИКИ, ПОКРЫТИЕ ПЛОСКОСТИ

Мозаикой называется всякое множество многоугольников, целиком покрывающее плоскость таким образом, что отдельные многоугольники не накрывают друг друга, и к каждой стороне произвольного многоугольника прилегает одна сторона соседнего многоугольника.

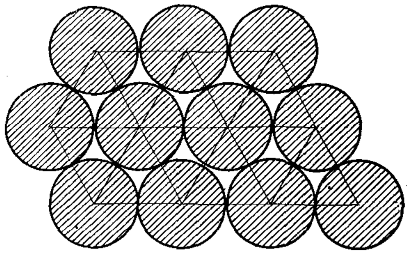


Мозаики, составленные из равных треугольников (a), из четырехугольников и шестиугольников (b), мозаика, составленная из пятиугольников (c)

До сегодняшнего дня остается невыясненным, какие вообще существуют мозаики. Поставим поэтому более конкретный вопрос: из каких равных выпуклых многоугольников можно составить мозаику? На этот вопрос уже можно дать гораздо более содержательный ответ. Очевидно, что мозаику можно составить из любых из равных параллелограммов, наиболее простой из них является квадратная мозаика — «клетчатая бумага». Поскольку всякий треугольник является половиной некоторого параллелограмма, то мозаику можно составлять и из любых треугольников. При составлении мозаики из произвольных четырехугольников следует симметрично отразить четырехугольник относительно середины одной из его сторон и прикладывать друг к другу полученные таким образом пары четырехугольников. Простейшей из треугольных мозаик является мозаика, составленная из правильных треугольников; из нее нетрудно получить шести- и пятиугольную мозаику. Можно доказать, что из равных выпуклых многоугольников с числом сторон, большим шести, уже нельзя составить мозаики. Если же отказаться от условия равенства многоугольников и допустить, например, лишь использование правильных многоугольников, то можно составлять самые разнообразные причудливые мозаики. На рисунке, например, изображена мозаика, составленная из четырех- и восьмиугольников.

Мозаики, составленные из 4- и 8-угольников

Равными кругами или другими равными выпуклыми фигурами, ограниченными кривыми, уже не удается покрыть плоскость без зазоров. В случае таких фигур интересно исследовать вопрос, как можно покрыть ими плоскость наиболее плотно. Точнее: как следует разместить выпуклые фигуры на плоскости таким образом, чтобы на произвольной части плоскости отношение площади покрытых участков к площади всей этой части было максимальным?



Наиболее плотное покрытие плоскости равными кругами

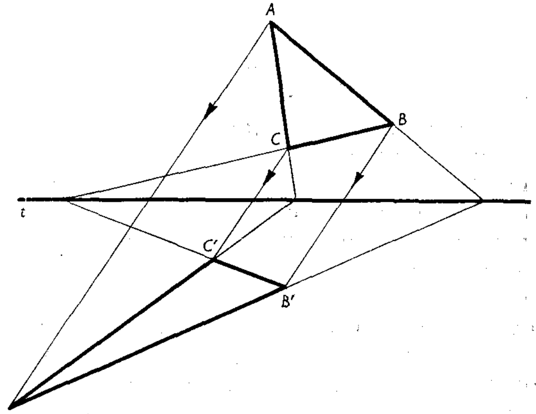
Это круг довольно сложных задач, очень многие из которых не решены до сих пор. Однако известен способ наиболее плотного размещения на плоскости кругов радиуса r: их центры должны быть расположены в вершинах мозаики, составленной из правильных треугольников со стороной $2r$. В связи с этим интересно отметить, что аналогичная задача наиболее плотного размещения равных шаров в пространстве пока не решена.

Вопросами такого рода занимается один из новых разделов геометрии: так называемая *дискретная геометрия*.

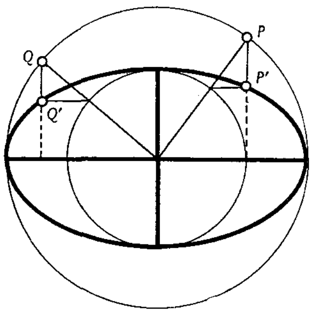
АФФИННЫЕ ПРЕОБРАЗОВАНИЯ

Одним из наиболее характерных свойств преобразований, сохраняющих расстояния, и преобразований подобия является тот факт, что при этих преобразованиях прямые переходят в прямые. Аналогичным свойством обладает, например, и проектирование плоскости на плоскость параллельно данному направлению; при этом каждой точке проектируемой плоскости ставится в соответствие одна из прямых некоторого пучка параллельных прямых; точка пересечения этой прямой со второй плоскостью и является образом исходной точки. Прямые пучка задают направление проектирования. Если они перпендикулярны плоскости образов, то говорят об ортогональной проекции. Преобразования, переводящие прямые в прямые, называются аффинными.

Как частные случаи, аффинные преобразования включают в себя преобразования, сохраняющие расстояния и преобразования подобия. При аффинном преобразовании сохраняется параллельность прямых. Отсюда, в частности, следует, что образом параллелограмма при аффинном преобразовании также является параллелограмм. Аффинные преобразования сохраняют отношения длин отрезков, лежащих на одной прямой или на параллельных прямых, а также отношение площадей. В то же



Треугольник и его образ при осевом аффинном преобразовании



Построение эллипса при помощи аффинного преобразования

время при аффинном преобразовании не обязательно сохраняются углы и длины отрезков.

Для любого аффинного преобразования существуют две взаимно перпендикулярные прямые, образы которых также являются взаимно перпендикулярными; направления, выбранные на этих прямых, называются главными направлениями аффинного преобразования.

Окружность переходит при аффинном преобразовании в эллипс (или окружность), оси которого лежат на главных направлениях.

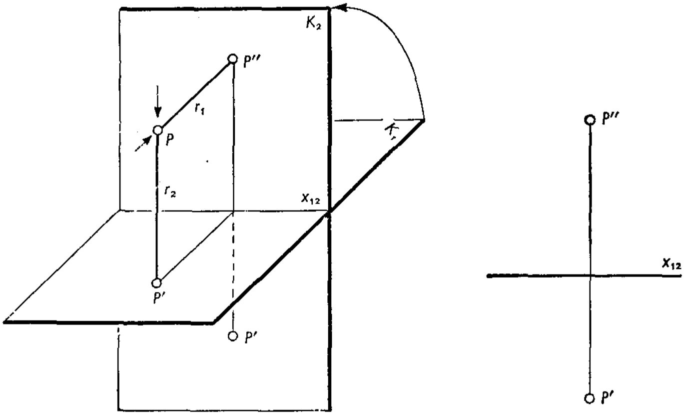
На практике чаще всего применяются аффинные преобразования плоскости на себя. Если в этом случае существует прямая, все точки которой являются неподвижными, то преобразование называется осевым аффинным преобразованием, а сама прямая — *осью* данного аффинного преобразования. При таком преобразовании все прямые, соединяющие точки плоскости с их образами, параллельны друг другу и задают направление аффинного преобразования. Всякая прямая пересекается своим образом в некоторой точке оси. Если направление преобразования перпендикулярно оси, то преобразование называется ортогональным. При ортогональном осевом аффинном преобразовании отношение расстояния произвольной точки от оси к расстоянию ее образа от оси есть величина постоянная, поэтому такое преобразование называют также аффинным сжатием.

Всякий эллипс можно рассматривать как образ его главной окружности при некотором аффинном сжатии, осью которого является большая ось эллипса; этот факт может быть использован и при построении эллипса.

НАЧЕРТАТЕЛЬНАЯ ГЕОМЕТРИЯ

Начертательная геометрия занимается методами построения изображений пространственных фигур на плоскости.

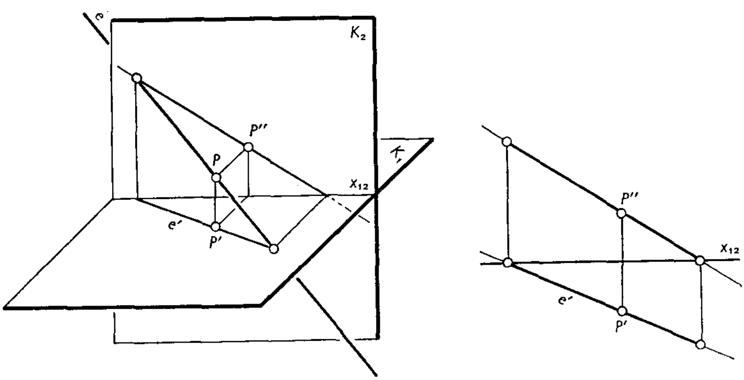
Для технических применений важное значение имеет задача изготовления таких чертежей технических объектов (например, деталей механизмов), по которым можно определить и восстановить вид исходного объекта и его положение в пространстве. Такое изображение всегда основано на принципе проекций. В зависимости от того, выполняется ли проектирование параллельно некоторому направлению или исходит из одной точки, говорят о параллельной или центральной проекции. Однако проекция не всегда осуществляет взаимно однозначное отображение, так как при проектировании пространственных тел, например, один и тот же образ может соответствовать многим точкам. Таким образом, по виду проекции не всегда удается восстановить вид тела однозначно. Различные методы построения изображений по-разному преодолевают эту трудность. Однозначно восстановить вид тела можно, например, с помощью



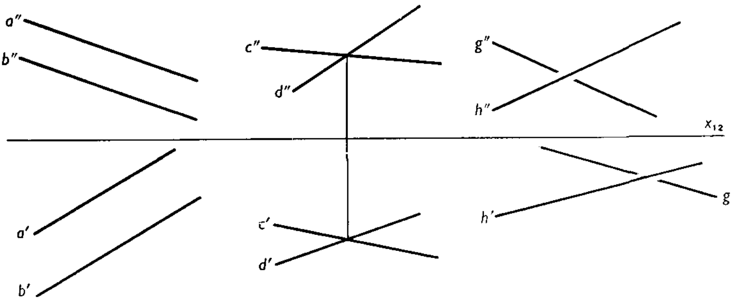
Изображение точки в двух проекциях

одновременного проектирования на несколько плоскостей. На этом основан *метод Монжа проекции на две плоскости* — один из старейших методов построения изображений. Суть его в том, что производится ортогональное проектирование на две взаимно перпендикулярные плоскости. С особым успехом этот метод применяется при решении задач на построение в пространстве. С его помощью пространственную задачу можно свести к задаче построения на плоскости, а затем перенести результат построения на исходный объект. После проектирования обе плоскости совмещают друг с другом, поворачивая их вокруг линии пересечения. Эти проекции называют, соответственно, *планом* (или первой проекцией) и *фасадом* (или второй проекцией).

В соответствии с этим мы будем говорить о *первой* (K_1) и *второй* (K_2) *плоскостях проекций*; линию их пересечения назовем осью x_{12}. При совмещении плоскостей проекций отрезки r_1 и r_2, соединяющие произвольную точку P с ее проекциями P' и P'', накладываются на прямую,



Изображение прямой e и ее точки P в проекциях на две плоскости



Пары параллельных, пересекающихся и скрещивающихся прямых

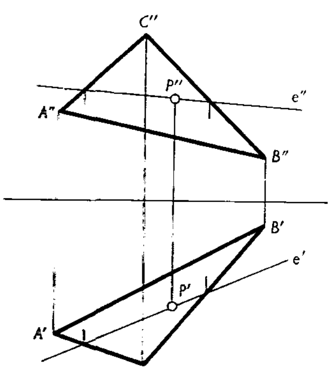
перпендикулярную оси. Таким образом две проекции какой-либо фигуры удовлетворяют тому условию, что прямые, соединяющие две проекции любой точки фигуры, перпендикулярны оси.

При выполнении построений ось, по которой пересекаются плоскости проекций, не играет существенной роли, поэтому, как правило, ее не рассматривают.

Если прямая не перпендикулярна ни одной из плоскостей проекций, то ее проекциями являются прямые; в противном случае одна из проекций есть точка.

Проекции параллельных прямых параллельны или совпадают; проекции пересекающихся прямых пересекаются. Если проекции двух прямых на обеих плоскостях пересекаются, то сами прямые могут или пересекаться (если прямая, соединяющая точки пересечения проекций в разных плоскостях, перпендикулярна оси) или быть скрещивающимися (в противном случае).

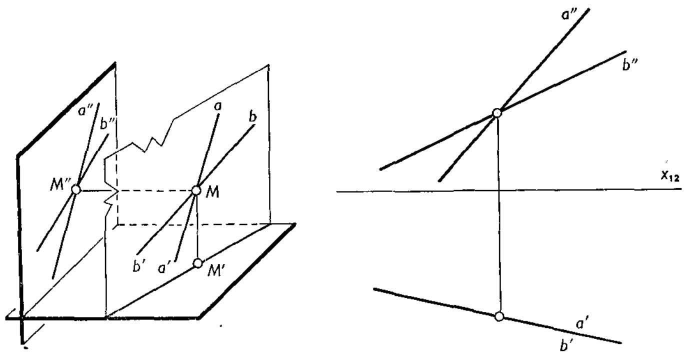
Плоскость можно задать с помощью двух прямых или одного треугольника. Прямая наверняка лежит в плоскости треугольника, если пересекает две из прямых, на которых лежат его стороны. Точка принадлежит плоскости, если она принадлежит некоторой прямой, лежащей в этой плоскости.



Задание плоскости с помощью треугольника. Изображение прямой и точки плоскости

Может случиться, что проекции двух прямых — скажем, на первую плоскость проекций — совпадают. Это может произойти, если прямые принадлежат плоскости, перпендикулярной к первой плоскости проекций (такая плоскость называется проецирующей). В самом деле, очевидно, что первая проекция всякой фигуры, лежащей в проецирующей плоскости, оказывается некоторой прямой или отрезком прямой.

Используя этот факт, можно построить *точку пересечения* прямой и плоскости. Зададим плоскость двумя проекциями некоторого па-

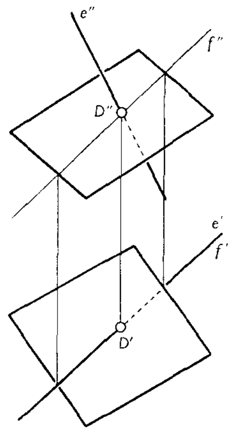


Первая плоскость проекций и прямые, проекции которых совпадают

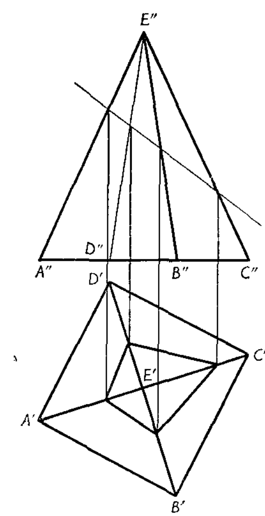
раллелограмма, принадлежащего ей. Найдем точку пересечения прямой e с этим параллелограммом. Рассмотрим на плоскости прямую f, первая проекция которой совпадает с первой проекцией прямой e. Обозначим через D'' точку пересечения прямых e'' и f'', являющихся вторыми проекциями прямых e и f. Тогда точка D'' является одновременно второй проекцией точки пересечения плоскости с прямой e, так как построенная таким образом точка D принадлежит и прямой e и плоскости параллелограмма (поскольку принадлежит прямой f, лежащей в этой плоскости).

Чтобы построить *линию пересечения* двух плоскостей, сначала выбирают на одной из них две прямые и находят точки их пересечения с другой плоскостью; поскольку эти точки принадлежат обеим плоскостями, то проходящая через них прямая и является искомой линией пересечения.

Нередко случается, что при соответствующем расположении фигуры относительно плоскостей проекций, задачи, связанные с пересечением, значительно упрощаются. Пусть, например, требуется построить линию пересечения четырехугольной пирамиды, основание которой принадлежит первой плоскости проекций, с некоторой плоскостью, перпендикулярной второй плоскости проекций. Построение значительно

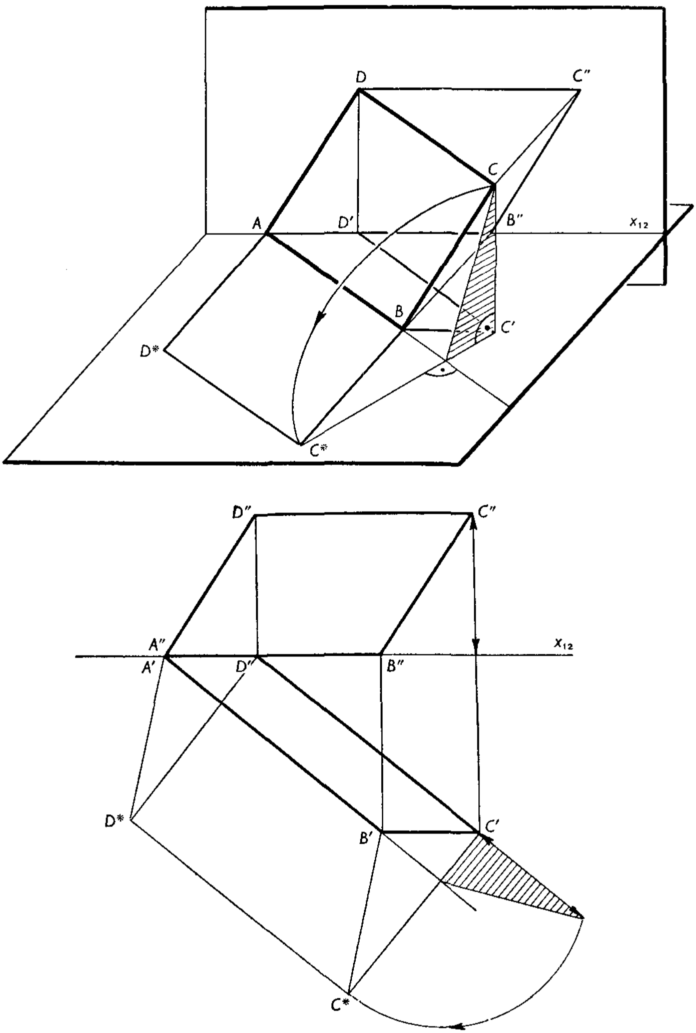


Точка пересечения прямой с плоскостью параллелограмма



Плоское сечение четырехугольной пирамиды упрощается, если заметить, что второй проекцией сечения является отрезок прямой. Таким образом точки пересечения плоскости с боковыми ребрами пирамиды могут быть непосредственно перенесены со второй проекции.

Более сложными оказываются, как правило, задачи, связанные с размерами объектов, например, определение размеров некоторой фигуры по ее проекциям или наоборот, построение проекций некоторой фигуры, если известны ее размеры. При решении таких задач фигуры стараются



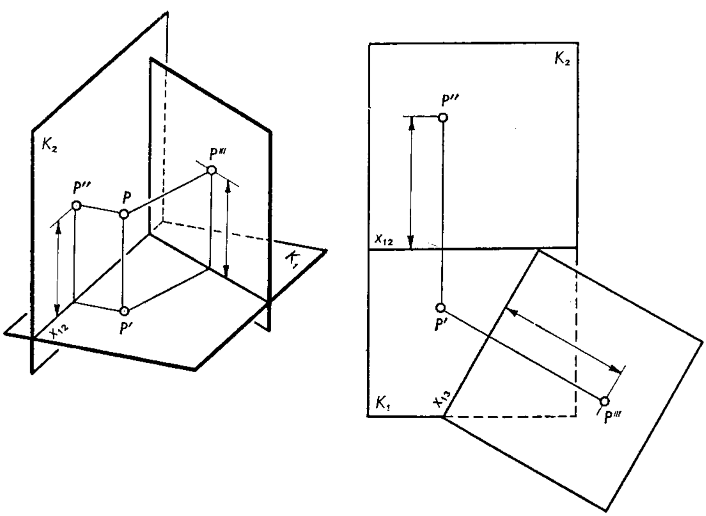
Определение истинных размеров параллелограмма с помощью поворота в плоскости проекций

расположить относительно плоскостей проекций таким образом, чтобы соотношения, связанные с размерами, можно было установить непосредственно. Плоская фигура, например, имеет на одной из проекций истинные размеры, если расположена в соответствующей плоскости проекций или параллельна ей.

В качестве примера рассмотрим параллелограмм *ABCD* (стр. 282), сторона *AB* которого лежит в первой плоскости проекций, а сторона *AD* — во второй. Если вращать параллелограмм вокруг стороны *AB*, то каждая его точка — и в том числе, например, вершина *C* — будет двигаться по некоторой окружности, центр которой лежит на стороне *AB*. Поскольку плоскость такой окружности перпендикулярна первой плоскости проекций, и оси вращения, то ее проекцией на первую плоскость будет некоторый отрезок, перпендикулярный отрезку *AB*. Следовательно, при вращении параллелограмма точка *C′* будет перемещаться вдоль некоторой прямой, перпендикулярной отрезку *AB*, до тех пор, пока точка *C* не достигнет первой плоскости проекций. Обозначим это положение точки *C* через *C**. Расстояние от точки *C** до оси вращения равно гипотенузе прямоугольного треугольника, один катет которого равен расстоянию от точки *C′* до оси вращения, а другой — расстоянию от точки *C* до первой плоскости проекций.

Полученный в результате вращения параллелограмм *ABC*D** имеет истинные размеры параллелограмма *ABCD*.

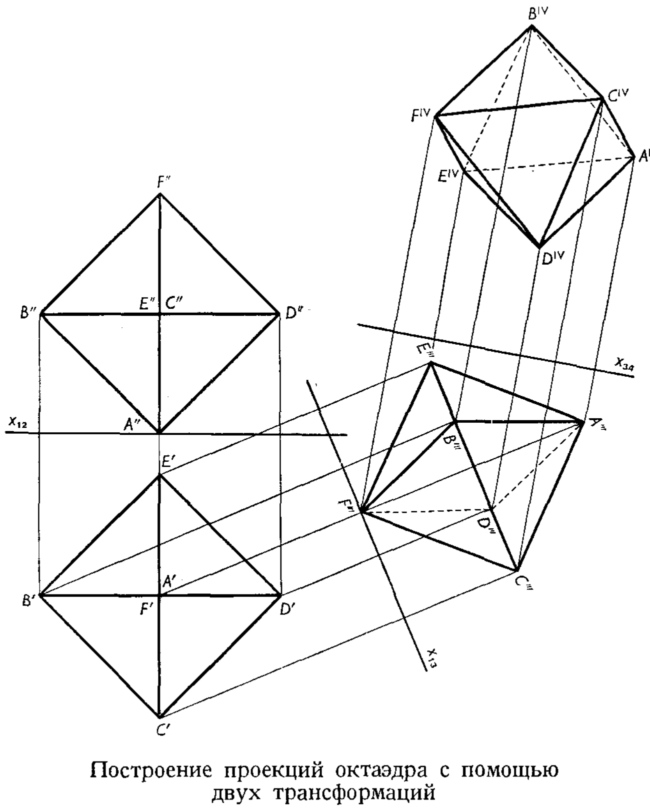
Заметим, что проекция плоской фигуры и ее изображение в той же плоскости, полученное в результате вращения, могут быть переведены друг в друга с помощью ортогонального аффинного преобразования. По этой причине для построения изображения «вращения» плоской фигуры достаточно выполнить его для одной из точек фигуры, тогда все изображение нетрудно построить, основываясь на свойствах аффинного преобразования.



Трансформация точки

Трансформация. Расположение фигуры или некоторой ее части относительно плоскостей проекций можно изменить и путем введения новых или дополнительных плоскостей проекций. Плоскость проекции, введенная дополнительно, должна быть перпендикулярна одной из «старых» плоскостей. Если, например, к плоскостям K_1 и K_2 добавляется плоскость K_3, перпендикулярная к K_1, то точка P''', являющаяся проекцией некоторой точки P на плоскость K_3, находится на таком же расстоянии от слоскости K_1, как и точка P''. Пусть мы хотим теперь исключить из рассмотрения плоскость K_2 и изобразить точку P в проекциях на плоскости K_1 и K_3, повернув для этого плоскость K_3 вокруг линии ее пересечения с плоскостью K_1 (вокруг оси x_{13}) до совмещения с K_1. Если ось x_{13} задана, то, согласно замечанию, сделанному выше, построить точку P''' не представляет труда: расстояние от нее до оси x_{13} равно расстоянию от точки P'' до оси x_{12}. Переход к новой системе плоскостей проекций принято называть *трансформацией* (см. рис. на стр. 283).

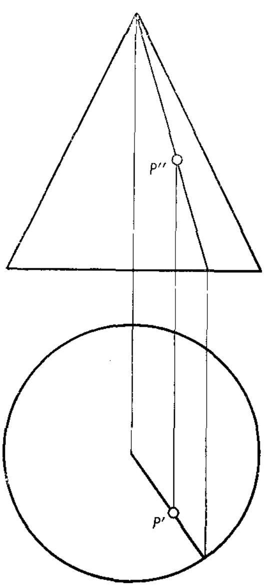
Трансформацию целесообразно применять и в тех случаях, когда положение фигуры относительно плоскостей проекций таково, что не дает достаточно наглядного представления о фигуре. На рисунке иозображе-



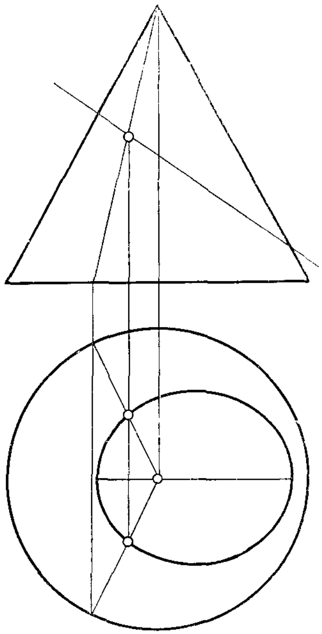
Построение проекций октаэдра с помощью двух трансформаций ны две проекции правильного тетраэдра и полученные из них с помощью трансформаций третья и четвертая проекции. Первые две проекции — квадраты с проведенными в них диагоналями. Основываясь на них, довольно трудно представить себе октаэдр. Значительно более наглядное представление об октаэдре дает четвертая проекция.

Построить точку на некоторой поверхности, например, на поверхности конуса вращения, можно с помощью линий, лежащих на этой поверхности. Для конуса вращения простейшими из таких линий являются его образующие. Линию пересечения конуса вращения с некоторой плоскостью можно построить, отметив точки пересечения произвольного числа образующих с этой плоскостью. На рисунке в качестве секущей плоскости для простоты выбрана вторая плоскость проекций; само сечение и его первая проекция представляют собой эллипс. Можно доказать, что один из фокусов эллипса, полученного в проекции, совпадает с проекцией вершины конуса.

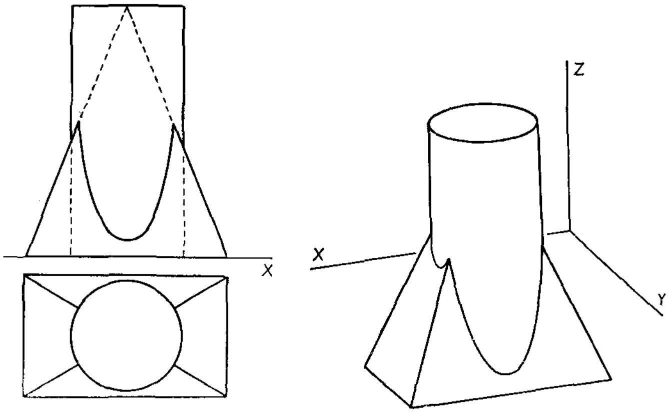
Аксонометрия. Метод изображения, называемый *аксонометрией,* применяется в первую очередь для построения наглядных чертежей. Суть его в следующем: изображаемый объект располагают в некоторой системе координат и с помощью параллельных лучей проектируют на некоторую плоскость. Геометрическое обоснование этого метода дается теоремой *Полке,* согласно которой три отрезка на плоскости, выходящие из одной точки, всегда могут рассматриваться как параллельные проек-



Изображение точки на поверхности конуса вращения



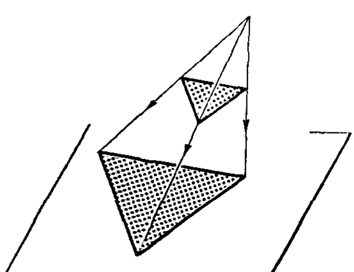
Построение плоского сечения конуса вращения



Изображение пирамиды и цилиндра по методу Монжа и с помощью аксонометрии

ции на эту плоскость трех ребер некоторого куба. На чертежах, построенных по методу Монжа или с помощью аксонометрии, сохраняются параллельность прямых и отношения пропорциональности (отношение отрезков, лежащих на одной прямой, например, деление пополам); плоская фигура и ее проекция могут быть переведены друг в друга при помощи аффинных преобразований, поэтому проекцией окружности, например, является эллипс. Проекции, построенные по методу Монжа, дают, как правило, лучшее представление о размерах объекта; за счет этого аксонометрия является более наглядной.

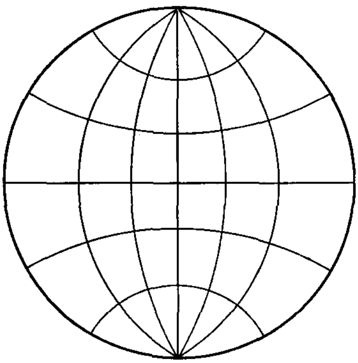
В случае *центральной проекции* обычно не сохраняются ни свойство параллельности, ни отношения пропорциональности. Центральная проекция является, например, геометрической основой фотографического изображения; существуют методы, позволяющие на основе фотографического снимка установить метрические свойства объекта (этим занимается *фотограмметрия*). Центральную проекцию называют также перспективой; перспективы параллельных прямых сходятся в одной точке, называемой *точкой схода*. Перспективой окружности может оказаться любое плоское сечение конуса. Обратно: всякое плоское сечение конуса может быть получено как центральная проекция окружности. Отсюда следует, что всякое плоское сечение конуса может быть получено в результате сечения некоторой плоскостью конуса вращения; всякое такое сечение можно рассматривать как центральную проекцию направляющей конуса вращения, то есть окружности.



Построение центральной проекции

Свойства центральных проекций (как правило, двух; с разными точками схода) широко используются в аэрофотосъемке и в ядерной физике («следы» треков частиц в пузырьковой камере и т. д.).

Стереографическая проекция является специальным видом центральной проекции; в этом случае поверхность шара проектируется из некоторой точки на шаре. Плоскость проекций должна быть параллельной плоскости, касательной к шару в точке, являющейся центром проекции. Этот метод проекции используется в первую очередь при изготовлении географических карт. Большое удобство представляет собой тот факт, что при стереографической проекции образами сферических окружностей являются окружности (или прямые). Кроме того эта проекция сохраняет углы между кривыми шаровой поверхности. Этим и объясняется тот факт, что на географических картах, изготовленных с помощью стереографической проекции окружности, сети координат (параллели и меридианы) изображаются дугами окружностей.



Проекция сети географических *координат. Центр лежит на* экваторе

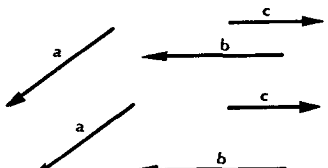
АНАЛИТИЧЕСКАЯ ГЕОМЕТРИЯ И ТРИГОНОМЕТРИЯ

ПОНЯТИЕ ВЕКТОРА

Сдвиг плоскости можно задать некоторым отрезком прямой. Для этого длина отрезка должна равняться расстоянию, на которое сдвигается плоскость, и отрезок должен быть направленным, то есть необходимо обозначить, в каком направлении (параллельно отрезку) происходит сдвиг. При изображении такого отрезка направление указывается стрелкой на его конце.

Направленный отрезок называется вектором.

Крайние точки называют, соответственно, началом и концом вектора. Вектор указывает направление от начала к концу. Вектор с началом в точке A и концом в точке B обозначается через $\overrightarrow{AB}$ (читается: вектор AB). Очень часто векторы обозначают одной буквой, напечатанной жирным шрифтом, например: **a**, **b**; иногда (например, в некоторых книгах по физике) для обозначения векторов используют готические буквы.



Одинаково обозначенные векторы равны

Один и тот же сдвиг плоскости можно задать с помощью бесконечного числа направленных отрезков. Такие отрезки имеют одинаковую длину и параллельны (или лежат на одной прямой). Векторы, задающие один и тот же сдвиг плоскости, принято считать равными, то есть два вектора равны, если они параллельны (или лежат на одной прямой), имеют одинаковую длину и направлены в одну сторону. Если векторы «параллельны или лежат на одной прямой», то они называются коллинеарными.

Длину вектора называют также его модулем, при этом модуль вектора **a** обозначается через |**a**|. Если модуль вектора равен 1, то вектор называется единичным. Используется также нулевой вектор, модуль которого равен 0, а направление считается произвольным. Нулевой вектор мы будем обозначать через **0** и изображать одной точкой.

СЛОЖЕНИЕ И ВЫЧИТАНИЕ ВЕКТОРОВ

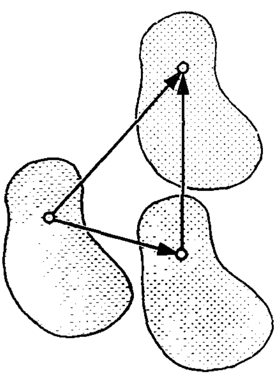
Как мы видели, два последовательных сдвига можно заменить одним, который называется суммой двух сдвигов. Зададим первый сдвиг вектором **a**, а второй — вектором **b**.

Если отложить вектор **b** от конца вектора **a**, то сумма двух сдвигов задается вектором, начало которого совпадает с началом вектора **a**, а конец — с концом вектора **b**. Этот вектор называется суммой векторов **a** и **b** и обозначается через **a** + **b**.

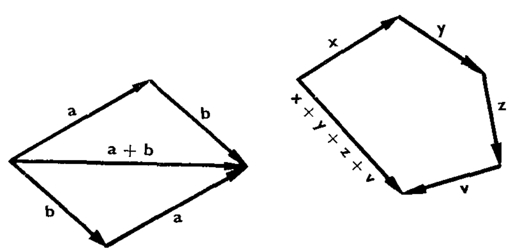
Сумму двух векторов можно построить и с помощью так называемого *правила параллелограмма*. Для этого оба вектора откладывают от одной точки и строят параллелограмм, сторонами которого являются эти векторы. Их суммой будет вектор, выходящий из исходной точки и совпадающий с соответствующей диагональю параллелограмма.

В физике, когда речь идет о силах, суммируемые векторы называют составляющими, а их сумму — равнодействующей.

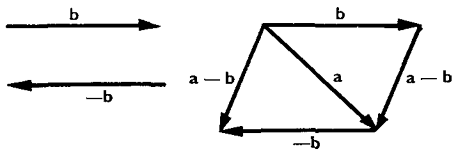
Согласно данному определению, чтобы получить сумму большего числа векторов, нужно отложить от произвольной точки первый вектор, а каждый последующий вектор отложить от конца предыдущего. Суммой будет



Сумма сдвигов, сумма векторов



Правило параллелограмма. Сумма нескольких векторов



Векторы $\mathbf{b}$ и $-\mathbf{b}$, вычитание векторов

вектор, начало которого совпадает с началом первого, а конец — с концом последнего вектора. Сложение векторов обладает теми же свойствами, что и сложение чисел: оно коммутативно ($\mathbf{a} + \mathbf{b} = \mathbf{b} + \mathbf{a}$) и ассоциативно $[(\mathbf{a} + \mathbf{b}) + \mathbf{c} = \mathbf{a} + (\mathbf{b} + \mathbf{c})]$.

Два вектора называются противоположными, если они коллинеарны и имеют один и тот же модуль, но направлены в противоположные стороны.

Сумма противоположных векторов равна 0. Вектор, противоположный вектору $\mathbf{b}$, обозначается через $-\mathbf{b}$. Под разностью между векторами $\mathbf{a}$ и $\mathbf{b}$ понимается сумма векторов $\mathbf{a}$ и $-\mathbf{b}$, которая обозначается через $\mathbf{a} - \mathbf{b}$. Это тот вектор, который при сложении с $\mathbf{b}$ дает $\mathbf{a}$.

УМНОЖЕНИЕ ВЕКТОРА НА ЧИСЛО

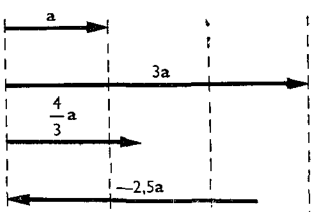
Следуя обозначению, принятому в алгебре, вектор $\mathbf{a} + \mathbf{a} + \mathbf{a}$ можно обозначить через $3\mathbf{a}$. Длина вектора $3\mathbf{a}$ втрое превышает длину вектора $\mathbf{a}$, они коллинеарны и направлены в одну сторону. Аналогично определяется и произведение вектора на произвольное число λ.

Под произведением $\lambda\mathbf{a}$ вектора $\mathbf{a}$ на число λ понимается такой вектор, который коллинеарен вектору $\mathbf{a}$, имеет модуль $|\lambda|\,|\mathbf{a}|$ и направлен в ту же сторону, что и $\mathbf{a}$, если λ положительно, и в противоположную — если λ отрицательно.

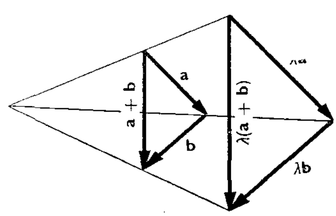
Геометрически умножение вектора на число означает растяжение или сжатие вектора и, возможно, перемену его направления на противоположное.

Умножение вектора на число удовлетворяет следующим важным свойствам:

$$\boxed{\begin{aligned}(\lambda + \mu)\mathbf{a} &= \lambda\mathbf{a} + \mu\mathbf{a},\\ \lambda(\mathbf{a} + \mathbf{b}) &= \lambda\mathbf{a} + \lambda\mathbf{b}.\end{aligned}}$$



Умножение вектора на число



$\lambda(\mathbf{a} + \mathbf{b}) = \lambda\mathbf{a} + \lambda\mathbf{b}$

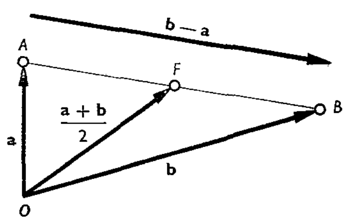
Отметим также, что произведение вектора на -1 равно противоположному вектору; два вектора коллинеарны тогда и только тогда, когда один из них равен произведению другого на некоторое число. Вместо $\frac{1}{\lambda}\,\mathbf{a}$ часто пишут $\frac{\mathbf{a}}{\lambda}$.

РАДИУС-ВЕКТОР

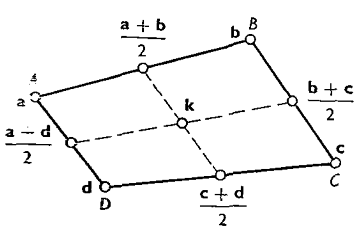
Из определения равенства векторов следует, что, задавая некоторый вектор, мы еще не знаем его положения в пространстве, так как задание одного вектора одновременно означает задание бесконечного числа равных ему векторов, «заполняющих» все пространство. По этой причине целесообразно сначала выбрать в пространстве некоторую фиксированную, так называемую *начальную точку* (обозначаемую, как правило, через O). Тогда произвольную точку P пространства можно однозначно задать вектором $\mathbf{p} = \overrightarrow{OP}$, начальной точкой которого является точка O, а конечной — точка P. В этом случае вектор $\mathbf{p}$ называется *радиус-вектором* точки P. Радиус-вектором точки O является нулевой вектор.

Если $\mathbf{a}$ — радиус-вектор точки A, а $\mathbf{b}$ — радиус-вектор точки B, то $\overrightarrow{AB} = \mathbf{a} - \mathbf{b}$, а радиус-вектором середины отрезка AB является вектор $\frac{\mathbf{a}+\mathbf{b}}{2}$. Вообще: радиус-вектором точки P, делящей отрезок AB в отношении $\lambda:\mu$ является вектор $\frac{\lambda\mathbf{a}+\mu\mathbf{b}}{\lambda+\mu}$.

С помощью этих результатов можно показать, что если $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$ — радиус-векторы вершин некоторого треугольника, то радиус-вектором его центра тяжести является вектор $\frac{\mathbf{a}+\mathbf{b}+\mathbf{c}}{3}$.



Радиус-векторы середин отрезка



Средние линии четырехугольника в точке пересечения делятся пополам

С помощью векторов нетрудно убедиться и в справедливости многих элементарных геометрических соотношений. Покажем, например, что средние линии некоторого четырехугольника $ABCD$ в точке своего пересечения делятся пополам. Пусть $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$, $\mathbf{d}$ — радиус-векторы вершин четырехугольника. Тогда радиус-векторами середин отрезков AB, BC, CD и DA будут, соответственно, $\frac{\mathbf{a}+\mathbf{b}}{2}$, $\frac{\mathbf{b}+\mathbf{c}}{2}$, $\frac{\mathbf{c}+\mathbf{d}}{2}$ и $\frac{\mathbf{d}+\mathbf{a}}{2}$. Радиус-вектором середины отрезка, соединяющего середины противоположных сторон AB и CD будет вектор

$$\mathbf{k} = \frac{\frac{\mathbf{a}+\mathbf{b}}{2} + \frac{\mathbf{c}+\mathbf{d}}{2}}{2} = \frac{\mathbf{a}+\mathbf{b}+\mathbf{c}+\mathbf{d}}{4}.$$

Аналогично получаем, что радиус-вектор середины отрезка, соединяющего середины сторон BC и DA равен

$$\frac{\frac{\mathbf{b}+\mathbf{c}}{2}+\frac{\mathbf{d}+\mathbf{a}}{2}}{2}=\frac{\mathbf{a}+\mathbf{b}+\mathbf{c}+\mathbf{d}}{4}.$$

Таким образом середины обеих средних линий имеют один и тот же радиус-вектор, а следовательно, совпадают (эту общую точку принято называть центром тяжести четырехугольника). (На рисунке мы не изобразили радиус-векторы).

РАЗЛОЖЕНИЕ ВЕКТОРА НА КОМПОНЕНТЫ; КООРДИНАТЫ ВЕКТОРА

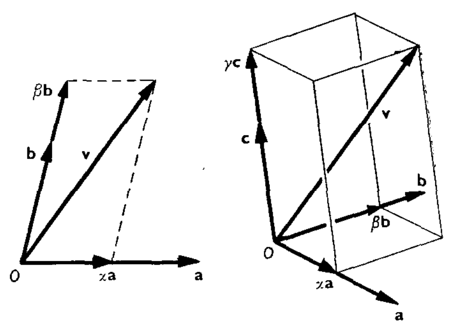
Пусть $\mathbf{a}$ и $\mathbf{b}$ — два неколлинеарных вектора в плоскости $п$, а $\mathbf{v}$ — произвольный вектор, лежащий в плоскости $п$ или параллельный ей. Отложим все три вектора от одной точки.

Для векторов $\mathbf{a}$ и $\mathbf{b}$ всегда найдутся такие числа α и β, что $\alpha\mathbf{a}+\beta\mathbf{b}=\mathbf{v}$. В этом случае говорят, что вектор $\mathbf{v}$ разложен по векторам $\mathbf{a}$ и $\mathbf{b}$.

Аналогично, если задать три вектора $\mathbf{a}$, $\mathbf{b}$ и $\mathbf{c}$, которые не параллельны одной плоскости, то произвольный вектор $\mathbf{v}$ пространства можно разложить по векторам $\mathbf{a}$, $\mathbf{b}$ и $\mathbf{c}$ следующим образом:

$$\mathbf{v}=\alpha\mathbf{a}+\beta\mathbf{b}+\gamma\mathbf{c}.$$

Заметим, что в обоих случаях разложение однозначно, то есть произвольный вектор может быть разложен по данным векторам единственным образом.



Разложение вектора на компоненты

Координаты. Если $\mathbf{i}$, $\mathbf{j}$, $\mathbf{k}$ — единичные векторы, не параллельные одной плоскости, и произвольный вектор $\mathbf{v}$ разлагается по этим векторам в виде

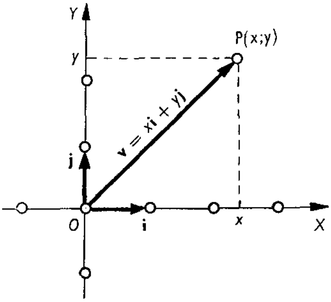
$$\mathbf{v}=x\mathbf{i}+y\mathbf{j}+z\mathbf{k},$$

то говорят, что вектор $\mathbf{v}$ имеет в системе $\mathbf{i}$, $\mathbf{j}$, $\mathbf{k}$ координаты (x, y, z).

Аналогично, если $\mathbf{i}$ и $\mathbf{j}$ — неколлинеарные векторы, и $\mathbf{v}$ — произвольный вектор на плоскости, который разлагается по векторам $\mathbf{i}$ и $\mathbf{j}$ в виде

$$\mathbf{v}=x\mathbf{i}+y\mathbf{j},$$

то числа (x, y) называются координатами вектора $\mathbf{v}$ в системе $\mathbf{i}$, $\mathbf{j}$.



Декартовы прямоугольные координаты на плоскости

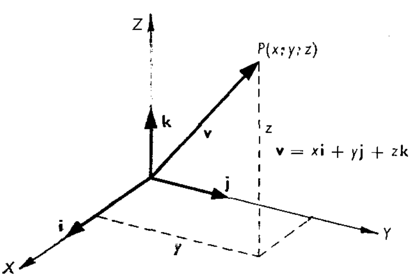
Если в последнем случае (на плоскости) векторы $\mathbf{i}$ и $\mathbf{j}$ взаимно ортогональны, причем — по принятому соглашению — вектор $\mathbf{j}$ может быть совмещен с $\mathbf{i}$ поворотом вправо, то говорят, что прямые, на которых лежат векторы $\mathbf{i}$ и $\mathbf{j}$, образуют декартову прямоугольную систему координат. В этом случае числа (x, y), присутствующие в разложении

$$\mathbf{v} = x\mathbf{i} + y\mathbf{j},$$

называются координатами вектора $\mathbf{v}$ в декартовой прямоугольной системе координат или просто декартовыми координатами вектора $\mathbf{v}$.

Упомянутые прямые называются осями координат. Таким образом, любому вектору соответствует некоторая пара чисел, и наоборот: всякой паре чисел соответствует некоторый вектор. Если отложить векторы $\mathbf{i}$, $\mathbf{j}$ и $\mathbf{v}$ от начальной точки O, то концу вектора $\mathbf{v}$ однозначно соответствует некоторая точка P плоскости.

Числа (x, y) называются также координатами точки P. Таким образом, координаты каждой точки совпадают с координатами ее радиус-вектора. Координатам точки можно придать простой геометрический смысл: они совпадают с длинами соответствующих перпендикуляров (с учетом знака), опущенных из точки на оси координат. (При этом длина перпендикуляра считается положительной, если он опускается на ось с той стороны, в которую указывает единичный вектор ($\mathbf{i}$ или $\mathbf{j}$) параллельной с ним оси, и отрицательной — в противном случае). По существу, оси X (на которой лежит вектор $\mathbf{i}$) и Y (на которой лежит вектор $\mathbf{j}$) — это прямые, пересекающиеся в точке O, а координаты произвольного вектора $\mathbf{v}$ могут быть получены как его проекции на оси X и Y (с учетом знака). Координаты некоторой точки P проще всего можно определить следующим образом: нужно опустить из этой точки перпендикуляры на оси координат, тогда числа, соответствующие основаниям перпендикуляров на осях (рассматриваемых как числовые оси) и будут координатами точки. Так, например, вторая координата точек, лежащих на оси X, и первая координата точек, лежащих на оси Y, всегда равны 0; координатами начала координат являются (0, 0).



Декартовы прямоугольные координаты в пространстве

Аналогично, если в пространстве выбрать единичные векторы $\mathbf{i}$, $\mathbf{j}$, $\mathbf{k}$ таким образом, чтобы они были попарно ортогональны и образовывали правую тройку (то есть располагались бы в пространстве так, как, соответственно, большой, указательный и средний пальцы правой руки), то говорят, что прямые, на которых лежат векторы, и проходящие через них плоскости образуют декартову прямоугольную систему координат в пространстве. Прямые, на которых лежат векторы $\mathbf{i}$, $\mathbf{j}$ и $\mathbf{k}$, называются, соответственно, осями X, Y и Z, а проходящие через них плоскости — координатными плоскостями. Числа (x, y, z), присутствующие в разложении

$$\mathbf{v} = x\mathbf{i} + y\mathbf{j} + z\mathbf{k},$$

называются прямоугольными координатами вектора $\mathbf{v}$. Их значения совпадают с расстояниями от конца вектора $\mathbf{v}$ до координатных плоскостей (с учетом знака).

Векторы $\mathbf{i}$, $\mathbf{j}$, $\mathbf{k}$ имеют, соответственно, координаты $(1, 0, 0)$, $(0, 1, 0)$, $(0, 0, 1)$.

Пусть вектор $\mathbf{a}$ имеет координаты (a_1, a_2, a_3), что обозначается также как $\mathbf{a}(a_1, a_2, a_3)$, и рассмотрим вектор $\mathbf{b}(b_2, b_2, b_3)$. Согласно сказанному выше,

$$\mathbf{a} = a_1\mathbf{i} + a_2\mathbf{j} + a_3\mathbf{k}$$

и

$$\mathbf{b} = b_1\mathbf{i} + b_2\mathbf{j} + b_3\mathbf{k}.$$

Складывая эти равенства, получаем:

$$\boxed{\mathbf{a} + \mathbf{b} = (a_1 + b_1)\mathbf{i} + (a_2 + b_2)\mathbf{j} + (a_3 + b_3)\mathbf{k}\,.}$$

Это значит, что вектор $\mathbf{a} + \mathbf{b}$ имеет координаты $(a_1 + b_1, a_2 + b_2, a_3 + b_3)$, то есть при сложении векторов их координаты складываются. Аналогичное утверждение справедливо и для разности векторов. При умножении вектора на число его координаты умножаются на то же число.

Рассмотрим, например, следующие три вектора:

$$\mathbf{a}(4, -3, 5),\ \mathbf{b}(-7, 9, 2),\ \mathbf{c}(2, 1, 0).$$

$\mathbf{a} + \mathbf{b}$ имеет координаты $(-3, 6, 7)$,

$\mathbf{a} - \mathbf{b}$ имеет координаты $(11, -12, 3)$,

$5\mathbf{a}$ имеет координаты $(20, -15, 25)$,

$\dfrac{\mathbf{a} + \mathbf{b}}{2}$ имеет координаты $(-1{,}5;\ 3;\ 3{,}5)$.

Векторы $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$ как радиус-векторы определяют некоторый треугольник, центру тяжести которого соответствует следующий радиус-вектор:

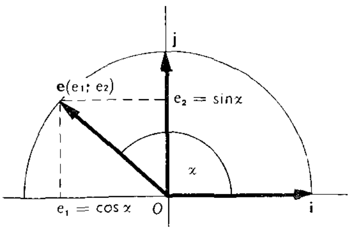
$$\frac{1}{3}(\mathbf{a} + \mathbf{b} + \mathbf{c})\left(-\frac{1}{3};\ \frac{7}{3};\ \frac{7}{3}\right).$$

Векторы $\mathbf{a}(4, -3, 5)$ и $\mathbf{d}(8, -6, 10)$ коллинеарны, так как из их координат видно, что $\mathbf{d} = 2\mathbf{a}$. Все эти правила, разумеется, имеют место и в системе координат на плоскости, только там векторы имеют по две координаты.

ОПРЕДЕЛЕНИЕ ТРИГОНОМЕТРИЧЕСКИХ ФУНКЦИЙ

Пусть **i** и **j** — единичные векторы некоторой системы координат на плоскости. Рассмотрим в этой плоскости произвольный единичный вектор **e**. Положение вектора **e** однозначно определяется заданием положительного угла α, на который следует повернуть вектор **i**, чтобы он совместился с **e**. Угол α мы назовем направляющим углом вектора **e**.

Задание направляющего угла однозначно определяет положение вектора **e**. С другой стороны, вектору **e** соответствует не один направляющий угол, так как углам α, $\alpha + 360°$, $\alpha + 2 \cdot 360°$, ..., $\alpha + k \cdot 360°$, ... — соответствует один и тот же вектор **e**.



Определение функций синус и косинус

Обозначим координаты вектора **e** через (e_1, e_2). Поскольку они однозначно определяются заданием направляющего угла α, то эти координаты являются функциями α. Эти функции получили специальное название: e_1 называется *косинусом* α, e_2 — *синусом* α. Обозначаются они следующим образом:

$$e_1 = \cos \alpha,$$

$$e_2 = \sin \alpha.$$

Косинус и синус некоторого угла α являются двумя координатами единичного вектора **e**, соответствующего углу α или, другими словами, обозначают длину направленных проекций единичного вектора на оси координат. Синус и косинус называются функциями угла или тригонометрическими функциями.

Принято вводить также следующие функции углов (которые определяются через уже известные функции):

тангенсом угла α ($\operatorname{tg} \alpha$) называется отношение синуса α к его косинусу (при условии, что $\cos \alpha \neq 0$); котангенсом угла α ($\operatorname{ctg} \alpha$) называется отношение косинуса α к его синусу (если $\sin \alpha \neq 0$).

$$\boxed{\operatorname{tg} \alpha = \frac{\sin \alpha}{\cos \alpha}, \qquad \operatorname{ctg} \alpha = \frac{\cos \alpha}{\sin \alpha}.}$$

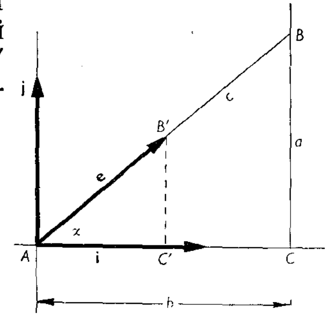
Реже используются функции *секанс* (sec) и *косеканс* (cosec):

$$\boxed{\sec \alpha = \frac{1}{\cos \alpha}, \qquad \operatorname{cosec} \alpha = \frac{1}{\sin \alpha}.}$$

Отметим, что для тангенса и котангенса иногда используются обозначения tan и cot (или cotan).

ФУНКЦИИ ОСТРОГО УГЛА И ПРЯМОУГОЛЬНЫЙ ТРЕУГОЛЬНИК

Расположим прямоугольный треугольник ABC в системе координат так, чтобы вершина A совпадала с началом координат, а катет AC лежал на оси X. Обозначим внутренний угол треугольника при вершине A через α, а стороны треугольника — через a, b и c. Рассмотрим единичный вектор **e**, соответствующий направляющему углу α и построим прямоугольный треугольник $AB'\,C'$, гипотенуза которого совпадает с вектором **e**, а катет AC' лежит на оси X. Рассматривая рисунок, нетрудно убедиться, что отношения сторон прямоугольного треугольника ABC могут быть выражены с помощью функций углов. Из подобия треугольников ABC и $AB'C'$ следует, что равны отношения их соответствующих сторон:



Функции острых углов

$$\boxed{\begin{array}{lll} \dfrac{BC}{AB} = \dfrac{B'C'}{AB'} & \text{т. е.} & \dfrac{a}{c} = \sin\alpha, \\[2ex] \dfrac{AC}{AB} = \dfrac{AC'}{AB'}, & \text{т. е.} & \dfrac{b}{c} = \cos\alpha. \end{array}}$$

Отсюда заключаем, что

$$\boxed{\frac{a}{b} = \operatorname{tg}\alpha \quad \text{и} \quad \frac{b}{a} = \operatorname{ctg}\alpha.}$$

Подведем итог: в прямоугольном треугольнике с острым углом α

синус угла α равен отношению катета, противолежащего этому углу, к гипотенузе;

косинус угла α равен отношению прилежащего катета к гипотенузе;

тангенс угла α равен отношению противолежащего катета к прилежащему;

котангенс угла α равен отношению прилежащего катета к противолежащему.

НЕКОТОРЫЕ ЗАМЕЧАНИЯ ОБ ОПРЕДЕЛЕНИИ ЗНАЧЕНИЙ ТРИГОНОМЕТРИЧЕСКИХ ФУНКЦИЙ

Значения тригонометрических функций определяют методами математического анализа и составляют таблицы этих значений. Однако таблицы содержат лишь значения функций острых углов, так как для других углов аналогичная задача может быть сведена к нахождению значений тригонометрических функций острых углов. Ниже мы приводим необходимые для этого соотношения. Пусть α — острый угол, тогда

$$\begin{array}{|rcl|}\hline \sin\alpha = \sin(180^\circ-\alpha) = -\sin(180^\circ+\alpha) = -\sin(360^\circ-\alpha), \\ \cos\alpha = -\cos(180^\circ-\alpha) = -\cos(180^\circ+\alpha) = \cos(360^\circ-\alpha), \\ \operatorname{tg}\alpha = -\operatorname{tg}(180^\circ-\alpha) = \operatorname{tg}(180^\circ+\alpha) = -\operatorname{tg}(360^\circ-\alpha), \\ \operatorname{ctg}\alpha = -\operatorname{ctg}(180^\circ-\alpha) = \operatorname{ctg}(180^\circ+\alpha) = -\operatorname{ctg}(360^\circ-\alpha). \\ \hline\end{array}$$

Если сумма двух углов равна 90° (обозначим один из них через α, а другой — через $90^\circ - \alpha$), то справедливы следующие соотношения:

$$\boxed{\begin{array}{ll} \sin\alpha = \cos(90^\circ-\alpha) & \cos\alpha = \sin(90^\circ-\alpha), \\ \operatorname{tg}\alpha = \operatorname{ctg}(90^\circ-\alpha) & \operatorname{ctg}\alpha = \operatorname{tg}(90^\circ-\alpha). \end{array}}$$

Тригонометрические функции являются *периодическими*. Это означает, что все их значения повторяются через равные промежутки. Такой минимальный промежуток, называемый *периодом*, для синуса и косинуса равен 360°, а для тангенса и котангенса — 180°, то есть

$$\boxed{\begin{array}{ll} \sin\alpha = \sin(\alpha + k\cdot 360^\circ), & \operatorname{tg}\alpha = \operatorname{tg}(\alpha + k\cdot 180^\circ), \\ \cos\alpha = \cos(\alpha + k\cdot 360^\circ), & \operatorname{ctg}\alpha = \operatorname{ctg}(\alpha + k\cdot 180^\circ). \\ k = (0, \pm 1, \pm 2, \ldots) & \end{array}}$$

Тригонометрические функции, естественным образом, определяются и для отрицательных углов. Это обеспечивают следующие соотношения:

$$\boxed{\begin{array}{ll} \sin(-\alpha) = -\sin\alpha, & \operatorname{tg}(-\alpha) = -\operatorname{tg}\alpha, \\ \cos(-\alpha) = \cos\alpha, & \operatorname{ctg}(-\alpha) = -\operatorname{ctg}\alpha. \end{array}}$$

Синус и косинус определены для любых углов. Тангенс (являющийся отношением синуса к косинусу) не имеет смысла там, где косинус равен 0, например, для 90° и 270°; аналогично, котангенс не определен для 0° и 180°.

Разделим область 0°—360° на четыре части, к которым отнесем, соответственно, углы 0°—90°, 90°—180°, 180°—270° и 270°—360°. В каждой из этих четвертей плоскости любая тригонометрическая функция имеет постоянный знак. Приводимая ниже таблица содержит необходимые сведения о знаках основных тригонометрических функций (через I, II, III и IV обозначены, соответственно, рассматриваемые четверти; знак «×» означает, что функция не определена).

	I	II	III	IV	0°	30°	45°	60°	90°	180°	270°
sin	+	+	−	−	0	$\frac{1}{2}$	$\frac{\sqrt{2}}{2}$	$\frac{\sqrt{3}}{2}$	1	0	−1
cos	+	−	−	+	1	$\frac{\sqrt{3}}{2}$	$\frac{\sqrt{2}}{2}$	$\frac{1}{2}$	0	−1	0
tg	+	−	+	−	0	$\frac{\sqrt{3}}{3}$	1	$\sqrt{3}$	×	0	×
ctg	+	−	+	−	×	$\sqrt{3}$	1	$\frac{\sqrt{3}}{3}$	0	×	0

Для тригонометрических функций одного аргумента выполняются также следующие соотношения:

$$\sin^2\alpha + \cos^2\alpha = 1,$$

$$\mathrm{tg}^2\,\alpha + 1 = \frac{1}{\cos^2\alpha} \qquad \mathrm{tg}\,\alpha\cdot\mathrm{ctg}\,\alpha = 1.$$

$$\mathrm{ctg}^2\,\alpha + 1 = \frac{1}{\sin^2\alpha}$$

ОСНОВНЫЕ ЗАДАЧИ ТРИГОНОМЕТРИИ

Задачей тригонометрии является определение сторон и углов треугольников, когда уже известны некоторые из них. Для этого в первую очередь используются тригонометрические функции. Причиной является то, что углы удается вычислять сравнительно проще и, как правило, точнее, чем расстояния. Кроме того, с помощью тригонометрии измерение расстояний часто может быть заменено измерением углов. С этой целью из соотношений между тригонометрическими функциями и длинами сторон прямоугольного треугольника можно вывести соотношения и теоремы, устанавливающие связь между длинами сторон и углами произвольного треугольника. Проще всего устанавливается зависимость между длинами сторон и углами прямоугольного треугольника. Обозначим его катеты через a и b, а гипотенузу — через c; пусть α обозначает угол, противоположный катету a, а β — угол, противоположный катету b. (Острый угол мы будем считать известным, если известно значение некоторой тригонометрической функции от него, так как в этом случае величину угла можно определить по таблице.)

1. Если известны a и b, то

$$\boxed{c = \sqrt{a^2 + b^2}, \qquad \operatorname{tg} \alpha = \frac{a}{b}, \qquad \beta = 90^\circ - \alpha.}$$

2. Если известны a и c, то

$$\boxed{b = \sqrt{c\ - a^2}, \qquad \sin \alpha = \frac{a}{c}, \qquad \beta = 90^\circ - \alpha.}$$

3. Если известны a и α, то

$$\boxed{b = a \operatorname{ctg} \alpha, \qquad c = \sqrt{a^2 + b^2}, \qquad \beta = 90^\circ - \alpha.}$$

4. Если известны c и α, то

$$\boxed{a = c \sin \alpha, \qquad b = c \cos \alpha, \qquad \beta = 90^\circ - \alpha.}$$

Для определения элементов произвольного треугольника чаще всего используются два соотношения: *теорема синусов* и *теорема косинусов*.

Пусть α, β и γ обозначают, соответственно, углы, противоположные сторонам a, b, и c треугольника. Имеют место следующие соотношения:

$$\boxed{\begin{array}{ll} a : b = \sin \alpha : \sin \beta & (\textit{теорема синусов}), \\ a^2 = b^2 + c^2 - 2bc \cos \alpha & (\textit{теорема косунусов}). \end{array}}$$

Эти соотношения, разумеется, останутся справедливыми, если заменить в них одну из сторон и противоположный ей угол, соответственно, другой стороной и противоположным ей углом треугольника.

С помощью теоремы синусов можно определить неизвестные элементы треугольника, когда известны две его стороны и угол, противоположный большей стороне, или когда дана одна сторона и два угла треугольника.

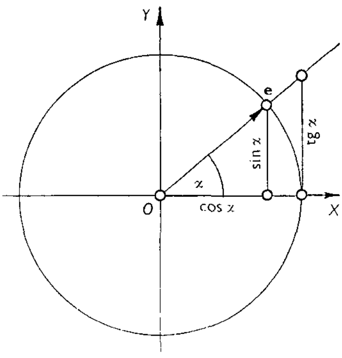
Теорема косинусов используется для нахождения неизвестных элементов, когда даны три стороны треугольника или две стороны и угол между ними.

Если известны две стороны (a, b) треугольника и угол (γ) между ними, то площадь t треугольника можно определить по формуле

$$\boxed{t = \frac{ab \sin \gamma}{2}.}$$

ТЕОРЕМЫ СЛОЖЕНИЯ И ИХ СЛЕДСТВИЯ

Значения тригонометрических функций можно находить и приближенно, так как, если в некоторой системе координат построить единичный вектор **e**, соответствующий направляющему углу α, то координаты вектора, то есть, согласно определению, синус и косинус α, можно измерить. Значение тангенса также может быть измерено, как длина отрезка касательной к единичной окружности от точки 1 оси X до пересечения с подвижным лучом (на котором лежит вектор **e**).



Функции углов на единичной окружности

Уже из рисунка можно заключить, что значения тригонометрических функций не возрастают пропорционально увеличению угла. Если исходный угол увеличить на некоторый угол, то, например, синус суммы двух углов не будет равен сумме синусов слагаемых, то есть

$$\sin(\alpha + \beta) \neq \sin \alpha + \sin \beta.$$

Имеют место следующие соотношения:

$$\boxed{\begin{aligned}
&\sin(\alpha+\beta) = \sin\alpha\cdot\cos\beta + \cos\alpha\cdot\sin\beta,\\
&\cos(\alpha+\beta) = \cos\alpha\cdot\cos\beta - \sin\alpha\cdot\sin\beta,\\
&\sin(\alpha-\beta) = \sin\alpha\cdot\cos\beta - \cos\alpha\cdot\sin\beta,\\
&\cos(\alpha-\beta) = \cos\alpha\cdot\cos\beta + \sin\alpha\cdot\sin\beta,\\
&\operatorname{tg}(\alpha+\beta) = \frac{\operatorname{tg}\alpha + \operatorname{tg}\beta}{1 - \operatorname{tg}\alpha\operatorname{tg}\beta}, \qquad \operatorname{tg}(\alpha-\beta) = \frac{\operatorname{tg}\alpha - \operatorname{tg}\beta}{1 + \operatorname{tg}\alpha\operatorname{tg}\beta}\\
&\operatorname{ctg}(\alpha+\beta) = \frac{\operatorname{ctg}\alpha\operatorname{ctg}\beta - 1}{\operatorname{ctg}\alpha + \operatorname{ctg}\beta}, \quad \operatorname{ctg}(\alpha-\beta) = \frac{\operatorname{ctg}\alpha\operatorname{ctg}\beta + 1}{\operatorname{ctg}\alpha - \operatorname{ctg}\beta}.
\end{aligned}}$$

Если в формулы сложения вместо β подставить α, то мы получим функции 2α:

$$\sin 2\alpha = 2 \sin \alpha \cos \alpha, \qquad \operatorname{tg} 2\alpha = \frac{2 \operatorname{tg} \alpha}{1 - \operatorname{tg}^2 \alpha},$$

$$\cos 2\alpha = \cos^2 \alpha - \sin^2 \alpha, \qquad \operatorname{ctg} 2\alpha = \frac{\operatorname{ctg}^2 \alpha - 1}{2 \operatorname{ctg} \alpha}.$$

Часто используются следующие формулы, которые также можно вывести из формул сложения:

$$\sin \frac{\alpha}{2} = \pm \sqrt{\frac{1 - \cos \alpha}{2}}, \qquad \cos \frac{\alpha}{2} = \pm \sqrt{\frac{1 + \cos \alpha}{2}}$$

$$\sin \alpha + \sin \beta = 2 \sin \frac{\alpha + \beta}{2} \cos \frac{\alpha - \beta}{2},$$

$$\sin \alpha - \sin \beta = 2 \cos \frac{\alpha + \beta}{2} \sin \frac{\alpha - \beta}{2},$$

$$\cos \alpha + \cos \beta = 2 \cos \frac{\alpha + \beta}{2} \cos \frac{\alpha - \beta}{2}$$

$$\cos \alpha - \cos \beta = -2 \sin \frac{\alpha + \beta}{2} \sin \frac{\alpha - \beta}{2}.$$

Если в тождества для $\sin 2\alpha$ и $\cos 2\alpha$ вместо α подставить $\frac{\alpha}{2}$, то получим, что

$$\sin \alpha = 2 \sin \frac{\alpha}{2} \cos \frac{\alpha}{2},$$

$$\cos \alpha = \cos^2 \frac{\alpha}{2} - \sin^2 \frac{\alpha}{2}.$$

Разделим каждое из этих равенств на тождество (почленно):

$$1 = \sin^2 \frac{\alpha}{2} + \cos^2 \frac{\alpha}{2}.$$

Сокращая дроби в правых частях на $\cos^2\frac{\alpha}{2}$ и вводя обозначение $\operatorname{tg}\frac{\alpha}{2}=t$, получаем:

$$\boxed{\begin{aligned}\sin\alpha &= \frac{2\sin\frac{\alpha}{2}\cos\frac{\alpha}{2}}{\sin^2\frac{\alpha}{2}+\cos^2\frac{\alpha}{2}} = \frac{2\operatorname{tg}\frac{\alpha}{2}}{\operatorname{tg}^2\frac{\alpha}{2}+1} = \frac{2t}{1+t^2},\\ \cos\alpha &= \frac{\cos^2\frac{\alpha}{2}-\sin^2\frac{\alpha}{2}}{\cos^2\frac{\alpha}{2}+\sin^2\frac{\alpha}{2}} = \frac{1-\operatorname{tg}^2\frac{\alpha}{2}}{1+\operatorname{tg}^2\frac{\alpha}{2}} = \frac{1-t^2}{1+t^2}.\end{aligned}}$$

Последние формулы показывают, что некоторые тригонометрические выражения можно заменять рациональными выражениями. Это используется в первую очередь при решении тригонометрических уравнений и в интегральном исчислении.

НЕКОТОРЫЕ НАИБОЛЕЕ ИЗВЕСТНЫЕ ТРИГОНОМЕТРИЧЕСКИЕ СООТНОШЕНИЯ

Если a и b — две стороны треугольника, а α и β — противоположные им углы, то, согласно теореме синусов,

$$\frac{a}{b}=\frac{\sin\alpha}{\sin\beta}.$$

Это можно переписать в виде:

$$\frac{a}{\sin\alpha}=\frac{b}{\sin\beta}.$$

Можно доказать, что величины $\frac{a}{\sin\alpha}$ и $\frac{b}{\sin\beta}$ равны диаметру d окружности, описанной около треугольника, то есть

$$\boxed{a=d\sin\alpha,\qquad b=d\sin\beta.}$$

Подставляя в дробь $\frac{a-b}{a+b}$ вместо a и b полученные выражения и сокращая на d, получаем:

$$\frac{a-b}{a+b}=\frac{d\sin\alpha-d\sin\beta}{d\sin\alpha+d\sin\beta}=\frac{\sin\alpha-\sin\beta}{\sin\alpha+\sin\beta}.$$

Подставим теперь в эту формулу выражения, полученные в предыдущем параграфе для $\sin\alpha-\sin\beta$ и $\sin\alpha+\sin\beta$.

$$\frac{a-b}{a+b}=\frac{2\cos\frac{\alpha+\beta}{2}\sin\frac{\alpha-\beta}{2}}{2\sin\frac{\alpha+\beta}{2}\cos\frac{\alpha-\beta}{2}}.$$

После упрощения приходим к следующему соотношению:

$$\boxed{\frac{a-b}{a+b}=\frac{\operatorname{tg}\frac{\alpha-\beta}{2}}{\operatorname{tg}\frac{\alpha+\beta}{2}},}$$

которое принято называть *теоремой тангенсов*.

Применение теоремы тангенсов позволяет иногда упростить вычисление неизвестных элементов треугольника. Если, например, даны стороны a и b треугольника и угол γ между ними, то углы α и β можно найти, используя соотношения

$$\frac{\alpha+\beta}{2}=\frac{180^\circ-\gamma}{2}=90^\circ-\frac{\gamma}{2}$$

и

$$\operatorname{tg}\frac{\alpha+\beta}{2}=\operatorname{tg}\left(90^\circ-\frac{\gamma}{2}\right)=\operatorname{ctg}\frac{\gamma}{2}.$$

Таким образом,

$$\operatorname{tg}\frac{\alpha-\beta}{2}=\frac{a-b}{a+b}\operatorname{ctg}\frac{\gamma}{2}.$$

Правая часть этого равенства содержит уже только известные данные, поэтому можно вычислить $\frac{\alpha-\beta}{2}$, то есть $\alpha-\beta$. С другой стороны, зная γ, мы знаем и $\alpha+\beta$.

Таким образом, искомые α и β могут быть найдены из тождеств

$$(\alpha + \beta) + (\alpha - \beta) = 2\alpha,$$
$$(\alpha + \beta) - (\alpha - \beta) = 2\beta.$$

Рассмотренную задачу можно решить и с помощью теоремы косинусов. Преимущество рассмотренного решения состоит в том, что при вычислениях приходится выполнять главным образом операции умножения и деления, а такие вычисления значительно упрощаются благодаря использованию логарифмов. Этим преимуществом обладают и приводимые ниже формулы для функций *половинного угла.*

Пусть a, b и c обозначают стороны треугольника, а α — угол, противоположный стороне a. Рассмотрим также периметр треугольника:

$$a + b + c = 2s.$$

Справедливы следующие соотношения:

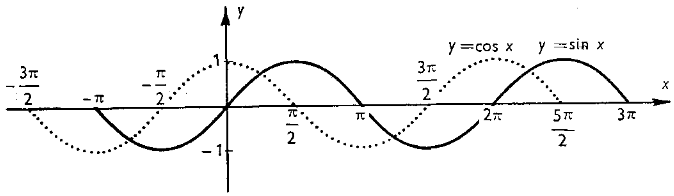
$$\boxed{\begin{aligned} \sin\frac{\alpha}{2} &= \sqrt{\frac{(s-b)(s-c)}{bc}}, \\ \cos\frac{\alpha}{2} &= \sqrt{\frac{s(s-a)}{bc}}, \\ \operatorname{tg}\frac{\alpha}{2} &= \sqrt{\frac{(s-b)(s-c)}{s(s-a)}}. \end{aligned}}$$

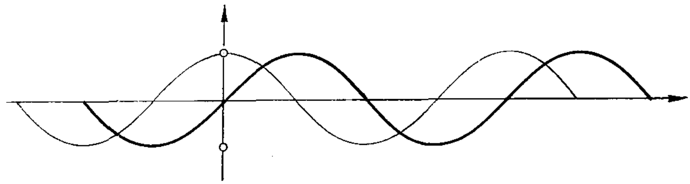
ГРАФИКИ ТРИГОНОМЕТРИЧЕСКИХ ФУНКЦИЙ

Изменение значений тригонометрических функций проще всего проследить, если построить их графики в прямоугольной системе координат. При изображении функции $y = \sin x$ переменная x означает угол, измеренный в радианах. Графиком этой функции является так называемая *синусоида* — волнообразная линия, длина «волны» которой равна 2π. Это объясняется тем, что значение функции при любом x совпадает с ее значением при $x + 2\pi$ (период функции $y = \sin x$ равен 2π).

График функции $y = \cos x$ получается из синусоиды $y = \sin x$ сдвигом на $\frac{\pi}{2}$ вдоль оси X в отрицательном направлении. Это следует из того, что

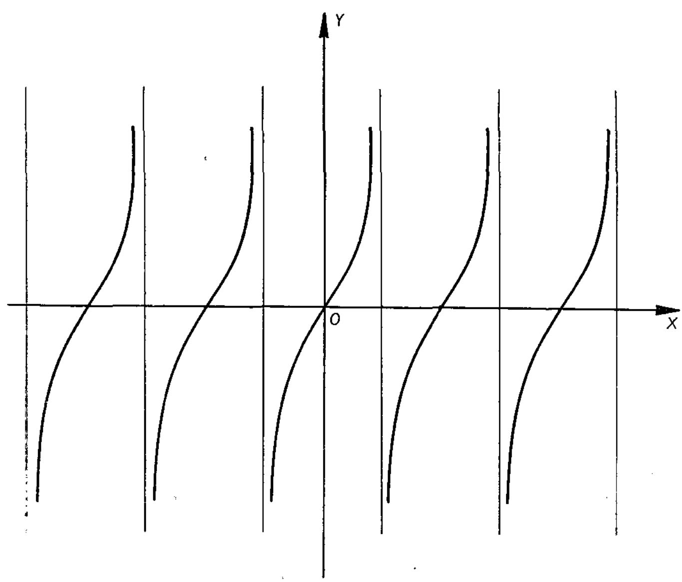
$$\cos x = \sin\left(x + \frac{\pi}{2}\right).$$



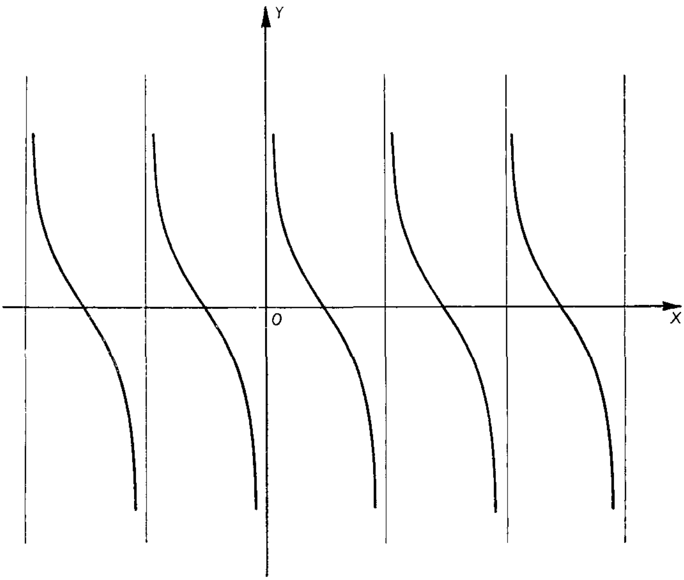


Графики функций $y = \sin x$ и $y = \cos x$

Достаточно точную синусоиду можно получить следующим образом: на предмет цилиндрической формы, сделанный из материала, легко поддающегося резанию (например, на свечку) плотно намотаем кусок бумажной ленты. После этого разрежем острым ножом бумагу вместе с цилиндром под углом 45° к образующей цилиндра. Размотав ленту, увидим, что обрезанный край имеет форму синусоиды.



Графики функции $y = \operatorname{tg} x$



Графики функции $y = \operatorname{ctg} x$

Функция $y = \operatorname{tg} x$ при стремлении x к $\frac{\pi}{2} \pm k\pi$ $(k = 0, 1, 2, \ldots)$ слева стремится к $+\infty$, а справа — к $-\infty$, поэтому ее график состоит из отдельных кривых, оба конца которых уходят в бесконечность. Тангенс имеет период π, поэтому при увеличении или уменьшении значения x на π значение тангенса не меняется.

График функции $y = \operatorname{ctg} x$ получается из графика $y = \operatorname{tg} x$ сдвигом на $\frac{\pi}{2}$ вдоль оси X и зеркальным отображением относительно оси X.

Величину угла α, синус которого равен z, называют *арксинусом* z и обозначают следующим образом:

$$\alpha = \arcsin z.$$

Функция $y = \arcsin x$ называется обратной функции $y = \sin x$. Аналогично определяются и обратные функции для $\cos x$, $\operatorname{tg} x$, $\operatorname{ctg} x$:

$$y = \arccos x,$$

$$y = \operatorname{arc\,tg} x,$$

$$y = \operatorname{arc\,ctg} x.$$

Поскольку одному значению $\sin x$ $(\cos x, \operatorname{tg} x, \operatorname{ctg} x)$ отвечает не одно x, то $\arcsin x$ $(\arccos x, \operatorname{arc\,tg} x, \operatorname{arc\,ctg} x)$ — неоднозначная функция. Поэтому часто под arc sin x, (arc cos x, arc tg x, arc ctg x) подразумевают лишь угол из интервала $\left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$ (соответственно, из интервалов однозначности): $[0, \pi]$, $\left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$ и $\left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$. Весь же спектр значений обозначают Arc sin x, (Arc cos x, Arc tg x, Arc ctg x).

ТРИГОНОМЕТРИЧЕСКИЕ УРАВНЕНИЯ

Уравнения, в которых неизвестное (или неизвестные) присутствует в качестве аргумента некоторой тригонометрической функции, называются *тригонометрическими уравнениями*.

К простейшим из них относится, например, уравнение

$$\operatorname{tg} x = 1.$$

Решить это уравнение — это значит, найти все те значения x, тангенс которых равен 1. Мы знаем, что $\operatorname{tg}\frac{\pi}{4} = \operatorname{tg} 45^\circ = 1$. Следовательно, $x = 45^\circ$ является решением этого уравнения, и между 0° и 180° других решений нет. Поскольку период тангенса равен 180°, то все решения могут быть заданыв виде:

$$x = 45^\circ \pm k \cdot 180^\circ \qquad (k = 0, 1, 2, \ldots)$$

Несколько сложнее решается уравнение

$$\sin x + \cos x = 1.$$

Трудность вызвана тем, что в уравнении присутствуют сразу две различные функции. Чтобы избавиться от этого, возведем обе части уравнения в квадрат:

$$\sin^2 x + \cos^2 x + 2 \sin x \cos x = 1.$$

Поскольку $\sin^2 x + \cos^2 x = 1$ и $2 \sin x \cos x = \sin 2x$, то после подстановки и упрощения уравнение приводится к виду:

$$\sin 2x = 0.$$

Отсюда получаем, что возможными значениями для $2x$ являются $\pm k \cdot 180^\circ$ ($k = 0, 1, 2, \ldots$). Следовательно,

$$x = \pm k \cdot 90^\circ.$$

В точках $x = \pm k \cdot 90^\circ$ синус и косинус могут принимать значения 0, 1 или -1. Сумма двух из этих чисел может равняться 1 только если одно равно 0, а другое 1. Между 0° и 360° это возможно лишь при 90°, поэтому общее решение уравнения имеет следующий вид:

$$x = 90° \pm k \cdot 360° \qquad (k = 0, 1, 2, \ldots)$$

Еще сложнее уравнение

$$(1 - \operatorname{tg} x)(1 + \sin 2x) = 1 + \operatorname{tg} x.$$

Подставим $\operatorname{tg} x = \dfrac{\sin x}{\cos x}$, а затем умножим обе части уравнения на $\cos x$:

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

$$(\cos x - \sin x)(1 + \sin 2x) = \cos x + \sin x.$$

Умножим обе части полученного уравнения на $\cos x + \sin x$, а затем произведем упрощения, учитывая, что $\cos^2 x + \sin^2 x = 1$, $\cos^2 x - \sin^2 x = \cos 2x$ и $2 \sin x \cos x = \sin 2x$:

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

или, преобразуя сумму в произведение:

$$(1 + \sin 2x)(\cos 2x - 1) = 0.$$

Поскольку произведение может равняться 0 только если хотя бы один из сомножителей равен 0, получаем, что возможными решениями являются

$$\begin{aligned} \sin 2x_1 &= -1, \\ 2x_1 &= 270° + k \cdot 360° \qquad (k = 0, \pm 1, \pm 2, \ldots) \\ x_1 &= 135° + k \cdot 180°, \end{aligned}$$

и

$$\begin{aligned} \cos 2x_2 &= 1, \\ 2x_2 &= k \cdot 360° \qquad (k = 0, \pm 1, \pm 2, \ldots) \\ x_2 &= k \cdot 180°. \end{aligned}$$

Полученные величины действительно удовлетворяют исходному уравнению.

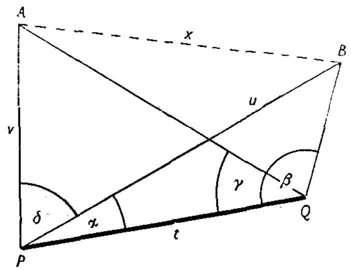
Уже из этих примеров видно, что нельзя дать общего метода решения тригонометрических уравнений, почти каждое из них требует особого подхода. Большинство таких уравнений вообще неразрешимо в привычном смысле, для них можно указать лишь метод приближенного решения·

Особенно часто такое положение возникает в случаях, когда, помимо тригонометрической функции, в уравнении присутствует какой-нибудь многочлен от неизвестного. Таково, например, уравнение

$$x = 2 \sin x.$$

ОБ ОДНОЙ ЗАДАЧЕ ИЗМЕРЕНИЯ НА МЕСТНОСТИ

Расстояние между двумя недоступными точками A и B на местности можно измерить, например, следующим образом: отмерим где-то на местности расстояние $PQ = t$, а затем с помощью соответствующего прибора (теодолита) измерим углы $\sphericalangle BPQ = \alpha$, $\sphericalangle BQP = \beta$, $\sphericalangle AQP = \gamma$ и $\sphericalangle APB = \delta$. На основании этих данных уже можно вычислить расстояние $x = AB$.



Определение расстояния AB

Из треугольника BPQ найдем сначала $\sphericalangle PBQ = \varepsilon$:

$$\varepsilon = 180^\circ - (\alpha + \beta)\,.$$

Применим теперь к треугольнику BPQ теорему синусов ($PB = u$):

$$u : t = \sin \beta : \sin \varepsilon,$$

$$u = \frac{t \sin \beta}{\sin \varepsilon}\,.$$

Из треугольника APQ можно найти $\sphericalangle PAQ = \omega$:

$$\omega = 180^\circ - (\alpha + \delta + \gamma)\,.$$

Теперь из треугольника APQ с помощью теоремы синусов определим длину $AP = v$:

$$v : t = \sin \gamma : \sin \omega,$$

$$v = \frac{t \sin \gamma}{\sin \omega}\,.$$

Таким образом в треугольнике ABP известны две стороны и угол между ними, следовательно, можно применить теорему косинусов:

$$x^2 = u^2 + v^2 - 2uv \cos \delta.$$

откуда:

$$x = \sqrt{u^2 + v^2 - 2uv \cos \delta}.$$

ТРИГОНОМЕТРИЯ СФЕРИЧЕСКИХ ТРЕУГОЛЬНИКОВ

Между сторонами и углами сферических треугольников существуют соотношения, аналогичные соотношениям для треугольников на плоскости. Напомним, что длины сторон сферических треугольников мы также измеряем углами: центральными углами, соответствующими их дугам. Пусть a, b и c — длины сторон сферического треугольника, а α, β и γ — противоположные им углы. Наиболее примечательны следующие соотношения между ними:

$$\sin a : \sin b = \sin \alpha : \sin \beta \qquad \textit{(теорема синусов)},$$

$$\cos c = \cos a \cos b + \sin a \sin b \cos \gamma$$

(теорема косинусов для сторон),

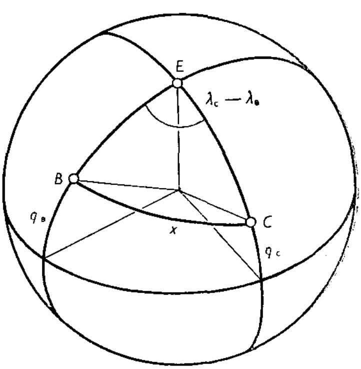
$$\cos \gamma = -\cos \alpha \cos \beta + \sin \alpha \sin \beta \cos c$$

(теорема косинусов для углов).

С помощью этих теорем по трем данным элементам сферического треугольника можно определить другие три (если задача вообще имеет решение). Эти теоремы используются в первую очередь при измерении расстояний на поверхности Земли, а также в задачах астрономии.

Теорема косинусов для сторон сферического треугольника позволяет определить расстояние между двумя точками земной поверхности, если известны их географические координаты. Если считать Землю шаром, то под географической широтой некоторого места земной поверхности следует понимать угол, который образует земной радиус, соответствующий этому месту, с плоскостью экватора, а под долготой — угол между начальным (нулевым) меридианом и меридианом, проходящим через данное место.

Географическая широта Будапешта (B) есть $\varphi_B = 47°30'$, а долгота $\lambda_B = 19°$. Соответствующие географические координаты Калькутты (C): $\varphi_C = 22°30'$, $\lambda_C = 88°22'$. Чтобы определить длину дуги, связывающей Будапешт с Калькуттой, рассмотрим сферический треугольник с вершинами в точках B, C и E (северный полюс). Имеем: $\sphericalangle BEC = \lambda_C - \lambda_B = 69°22'$; стороне BE соответствует угол $90° - \varphi_B = 42°30'$, а стороне CE — угол $90° - \varphi_C = 67°30'$. На основе этих данных по теореме синусов получаем:



Определение расстояния между Будапештом и Калькуттой

$$\cos x = \cos 42°30' \cdot \cos 67°30' +$$
$$+ \sin 42°30' \cdot \sin 67°30' \cdot \cos 69°22',$$

где x обозначает угол, соответствующий дуге BC. Отсюда для x получается приближенное значение $59°50'$. Поскольку дуга (большого радиуса), соответствующая центральному углу $1°$, имеет длину около 111 км, то расстояние от Будапешта до Калькутты приблизительно равно 6600 км.

СКАЛЯРНОЕ ПРОИЗВЕДЕНИЕ ВЕКТОРОВ

Во множестве векторов определяют произведения нескольких типов. Кроме уже рассмотренного умножения вектора на число, очень часто используется так называемое *скалярное умножение* векторов. Само название указывает на то, что в этом случае произведением является *скаляр*, то есть *число*.

Под скалярным произведением векторов понимается произведение их модулей на косинус угла между ними.

Таким образом, если через φ обозначить угол между векторами **a** и **b**, а через **ab** — их скалярное произведение, то

$$\boxed{\mathbf{ab} = |\mathbf{a}|\,|\mathbf{b}|\cos\varphi.}$$

Скалярное произведение обладает двумя обычными свойствами умножения: коммутативностью и дистрибутивностью, то есть

$$\boxed{\begin{array}{c} \mathbf{ab} = \mathbf{ba} \\ \text{и} \\ \mathbf{a}(\mathbf{b} + \mathbf{c}) = \mathbf{ab} + \mathbf{ac}. \end{array}}$$

В то же время ассоциативность не выполняется, то есть в общем случае

$$\mathbf{a}(\mathbf{bc}) \neq (\mathbf{ab})\mathbf{c}.$$

Если векторы взаимно ортогональны, то угол между ними равен $90°$, и посколько косинус $90°$ равен 0, то скалярное произведение ортогональных векторов равно 0. Обратно: если скалярное произведение двух векторов равно 0, то эти векторы взаимно ортогональны.

Скалярное произведение легко определяется, если известны координаты векторов: скалярное произведение двух векторов равно сумме произведений их соответственных координат. Например, скалярное произведение векторов $\mathbf{a}(a_1;\ a_2;\ a_3)$ и $\mathbf{b}(b_1;\ b_2;\ b_3)$ вычисляется по следующей формуле:

$$\boxed{\mathbf{ab} = a_1b_1 + a_2b_2 + a_3b_3\,.}$$

Если $\mathbf{a} = \mathbf{b}$, то косинус угла ($0°$) равен 1 и, следовательно,

$$\mathbf{a} \cdot \mathbf{a} = \mathbf{a}^2 = |\mathbf{a}|\,|\mathbf{a}| \cdot 1 = |\mathbf{a}|^2.$$

В координатной форме:

$$\boxed{\begin{aligned} |\mathbf{a}|^2 &= a_1^2 + a_2^2 + a_3^2, \\ |\mathbf{a}| &= \sqrt{a_1^2 + a_2^2 + a_3^2}. \end{aligned}}$$

Таким образом, модуль вектора равен квадратному корню из суммы квадратов его координат. Если разделить вектор на его модуль, то мы получим единичный вектор того же направления. Единичный вектор направления **a** имеет следующие координаты:

$$\left(\frac{a_1}{\sqrt{a_1^2 + a_2^2 + a_3^2}}; \quad \frac{a_2}{\sqrt{a_1^2 + a_2^2 + a_3^2}}; \quad \frac{a_3}{\sqrt{a_1^2 + a_2^2 + a_3^2}}\right).$$

Зная скалярное произведение, можно определить угол между векторами:

$$\boxed{\cos\varphi = \frac{\mathbf{ab}}{|\mathbf{a}| \cdot |\mathbf{b}|}.}$$

Поскольку косинус острого угла положителен, а косинус тупого — отрицателен, то из этой формулы вытекает, что если скалярное произведение положительно, то векторы образуют острый угол, а если отрицательно — тупой.

Рассмотрим, например, векторы $\mathbf{a}(4;\ -3;\ 5)$ и $\mathbf{b}(-2;\ -9;\ 4)$. Их скалярное произведение

$$\mathbf{ab} = -4 \cdot 2 + 3 \cdot 9 + 5 \cdot 4 = 39,$$

следовательно, эти векторы образуют острый угол. Вычислим их модули:

$$|\mathbf{a}| = \sqrt{16 + 9 + 25} = \sqrt{50} = 7{,}07,$$

$$|\mathbf{b}| = \sqrt{4 + 81 + 16} = \sqrt{101} = 10{,}05.$$

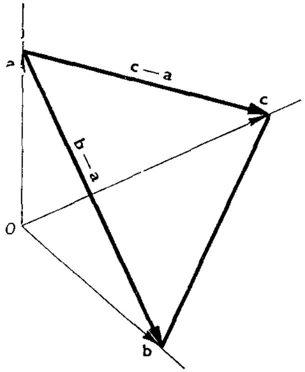
Единичный вектор направления **a** имеет координаты $\left(\frac{4}{\sqrt{50}};\ -\frac{3}{\sqrt{50}};\ \frac{5}{\sqrt{50}}\right)$.

Определим угол между **a** и **b**:

$$\cos\varphi = \frac{\mathbf{ab}}{|\mathbf{a}|\,|\mathbf{b}|} = \frac{39}{\sqrt{50} \cdot \sqrt{101}} = 0{,}5485,$$

$$\varphi = 56°44'.$$

Скалярное произведение векторов может быть использовано при решении различных задач и при доказательстве теорем.

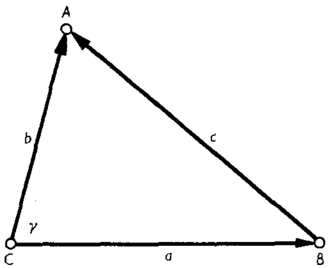


Треугольное сечение куба

Пример: Докажем, что если три ребра куба, выходящие из одной вершины, пересечь плоскостью, то в сечении получится треугольник, все углы которого острые. Выберем в качестве начала координат общую вершину и обозначим через **a**, **b** и **c** векторы, выходящие из начала координат, с концами в вершинах треугольника, полученного в результате сечения. Рассмотрим произвольный угол треугольника, например, угол у конца вектора **a**. Это угол между векторами **b** — **a** и **c** — **a**. Нужно показать, что скалярное произведение этих векторов положительно, так как это и означает, что рассматриваемый угол острый.

$$(\mathbf{b}-\mathbf{a})(\mathbf{c}-\mathbf{a})=\mathbf{bc}-\mathbf{ac}-\mathbf{ab}+\mathbf{a}^2.$$

Первые три слагаемые в правой части равны 0, так как являются произведениями взаимно ортогональных векторов, но $\mathbf{a}^2$ положительно и, следовательно, рассматриваемое скалярное произведение в самом деле положительно.



Доказательство теоремы косинусов с помощью векторов

Другой **пример:** Пусть a, b и c — три стороны треугольника ABC, а γ — угол, противоположный стороне c. Установим направление вдоль сторон таким образом, чтобы выполнялось равенство

$$\overrightarrow{CA}-\overrightarrow{CB}=\overrightarrow{BA}.$$

Возводя обе части этого равенства в квадрат и учитывая, что $|\overrightarrow{CB}|=a$, $|\overrightarrow{CA}|=b$ и $|\overrightarrow{BA}|=c$, получаем:

$$\overrightarrow{CA}^2+\overrightarrow{CB}^2-2\overrightarrow{CA}\cdot\overrightarrow{CB}=\overrightarrow{BA}^2,$$

$$a^2+b^2-2ab\cos\gamma=c^2,$$

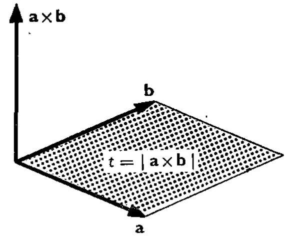
а это как раз теорема косинусов.

ВЕКТОРНОЕ ПРОИЗВЕДЕНИЕ ; СМЕШАННОЕ ПРОИЗВЕДЕНИЕ

Под векторным произведением векторов **a** и **b** — оно обозначается через $\mathbf{a}\times\mathbf{b}$ — понимается такой ортогональный векторам **a** и **b** вектор, что **a**, **b**, $\mathbf{a}\times\mathbf{b}$ образуют правую тройку, а модуль его равен $|\mathbf{a}|\,|\mathbf{b}|\sin\varphi$ (где φ — угол между векторами **a** и **b**).

Последнее условие означает, что если смотреть в направлении вектора $\mathbf{a}\times\mathbf{b}$, то $\mathbf{a}$ совмещается с $\mathbf{b}$ поворотом вправо.

Согласно определению векторного произведения, между направляющими векторами $\mathbf{i}$, $\mathbf{j}$ и $\mathbf{k}$ системы координат имеют место следующие соотношения:



Векторное произведение

$$\boxed{\mathbf{i}\times\mathbf{j}=\mathbf{k},\qquad \mathbf{j}\times\mathbf{k}=\mathbf{i},\qquad \mathbf{k}\times\mathbf{i}=\mathbf{j}.}$$

Векторное произведение уже не коммутативно; перемена сомножителей местами приводит к изменению направления вектора-произведения:

$$\boxed{\mathbf{a}\times\mathbf{b}=-\mathbf{b}\times\mathbf{a}.}$$

В то же время выполняется дистрибутивность:

$$\boxed{(\mathbf{a}+\mathbf{b})\times\mathbf{c}=\mathbf{a}\times\mathbf{c}+\mathbf{b}\times\mathbf{c}.}$$

Отметим геометрический смысл модуля вектора $\mathbf{a}\times\mathbf{b}$: он равен площади параллелограмма, образованного векторами $\mathbf{a}$ и $\mathbf{b}$. Если задать координаты векторов $\mathbf{a}$ и $\mathbf{b}$, то

$$\boxed{\mathbf{a}\times\mathbf{b}=\begin{vmatrix}\mathbf{i} & \mathbf{j} & \mathbf{k}\\ a_1 & a_2 & a_3\\ b_1 & b_2 & b_3\end{vmatrix}}$$

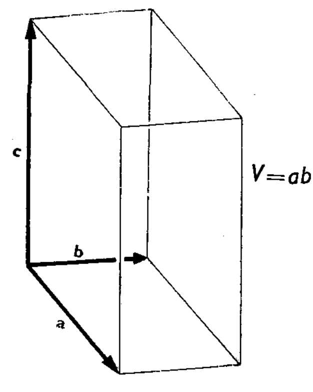
Причем, в разложении этого определителя координаты вектора $\mathbf{a}\times\mathbf{b}$ присутствуют как коэффициенты при векторах $\mathbf{i}$, $\mathbf{j}$ и $\mathbf{k}$.

Смешанное произведение.

Смешанным произведением трех векторов называется скалярное произведение векторного произведения первых двух векторов на третий вектор.

Смешанное произведение векторов $\mathbf{a}$, $\mathbf{b}$ и $\mathbf{c}$ обозначается через $\mathbf{abc}$. Таким образом, смешанное произведение есть число.

$$\boxed{\mathbf{abc}=(\mathbf{a}\times\mathbf{b})\mathbf{c}.}$$



Смешанное произведение

Смешанное произведение зависит от порядка сомножителей, точнее:

$$\boxed{\mathbf{abc} = \mathbf{bca} = \mathbf{cab} = -\mathbf{acb} = -\mathbf{cba} = -\mathbf{bac}.}$$

Геометрически смешанное произведение выражает величину объема параллелепипеда, построенного на трех перемножаемых векторах, взятого со знаком плюс или минус согласно тому, образуют ли векторы **a**, **b** и **c** правую или левую тройку. Смешанное произведение векторов $\mathbf{a}(a_1;\ a_2;\ a_3)$, $\mathbf{b}(b_1;\ b_2;\ b_3)$ и $\mathbf{c}(c_1;\ c_2;\ c_3)$ вычисляется следующим образом:

$$\boxed{\mathbf{abc} = \begin{vmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{vmatrix}.}$$

Смешанное произведение трех векторов равно 0 тогда и только тогда, когда все три вектора лежат в одной плоскости.

Часто используется следующее соотношение, в некоторой степени выражающее связь между векторным и скалярным произведениями:

$$\boxed{(\mathbf{a}\times\mathbf{b})\times\mathbf{c} = (\mathbf{ac})\mathbf{b} - (\mathbf{bc})\mathbf{a}} \qquad \text{(теорема разложения)}.$$

РАССТОЯНИЕ МЕЖДУ ДВУМЯ ТОЧКАМИ

Пусть две точки заданы своими координатами: $P_1(x_1;\ y_1;\ z_1)$ и $P_2(x_2;\ y_2;\ z_2)$. Это означает, что радиус-векторами точек P_1 и P_2 являются, соответственно, векторы $\mathbf{p}_1(x_1;\ y_2;\ z_1)$ и $\mathbf{p}_2(x_2;\ y_2;\ z_2)$. Рассмотрим вектор с началом в точке P_1 и концом в P_2:

$$\overrightarrow{P_1P_2} = \mathbf{p}_2 - \mathbf{p}_1(x_2 - x_1;\quad y_2 - y_1;\quad z_2 - z_1).$$

Расстояние P_1P_2 равно модулю вектора $\mathbf{p}_2 - \mathbf{p}_1$, т. е.

$$\boxed{P_1P_2 = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2 + (z_1 - z_2)^2}.}$$

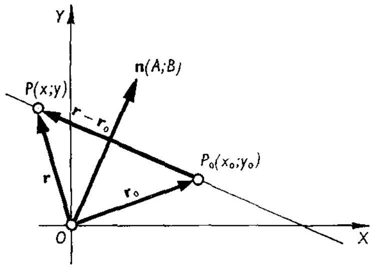
Если точки P_1 и P_2 заданы своими координатами в некоторой прямоугольной системе координат на плоскости, то третья координата равна 0, следовательно,

$$\boxed{P_1P_2 = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}.}$$

АНАЛИТИЧЕСКАЯ ГЕОМЕТРИЯ ПРЯМОЙ

Зададим на плоскости систему координат и рассмотрим в этой плоскости произвольную прямую. Положение прямой определено однозначно, если известны некоторая ее точка и вектор, перпендикулярный прямой — так называемый *вектор нормали* к прямой.

Пусть $\mathbf{r}_0(x_0; y_0)$ — радиус-вектор данной точки P_0 прямой, а $\mathbf{n}(A; B)$ — вектор нормали к прямой. Обозначим через $\mathbf{r}(x; y)$ радиус-вектор произвольной точки P' прямой.



Уравнение прямой

Вектор $\mathbf{r} - \mathbf{r}_0$ параллелен прямой, следовательно, ортогонален вектору $\mathbf{n}$, поэтому их скалярное произведение равно 0:

$$\mathbf{n}(\mathbf{r} - \mathbf{r}_0) = 0.$$

В этом уравнении $\mathbf{n}$ и $\mathbf{r}_0$ фиксированы, а конец вектора $\mathbf{r}$ пробегает всю прямую, причем равенство выполняется при любом его положении. Верно и обратное утверждение: если некоторый вектор $\mathbf{r}$ удовлетворяет уравнению $\mathbf{n}(\mathbf{r} - \mathbf{r}_0) = 0$, то он является радиус-вектором некоторой точки данной прямой. Таким образом это, уравнение задает прямую и называется *векторным уравнением прямой*. Переписывая его в координатной форме, получаем следующее *уравнение прямой*:

$$\boxed{A(x - x_0) + B(y - y_0) = 0.}$$

Это уравнение можно переписать в виде:

$$\boxed{Ax + By = Ax_0 + By_0.}$$

Постоянную величину в правой части принято обозначать одной буквой, например, C;

$$\boxed{Ax + By = C.}$$

Итак, уравнением прямой мы называем такое уравнение, которому удовлетворяют координаты всех точек этой прямой и только они.

Уравнение прямой есть уравнение первой степени с двумя неизвестными.

Если прямая параллельна оси X или оси Y, то, соответственно, A или B равняется нулю. Если же прямая пересекает ось X в точке $(a; 0)$, а ось Y — в точке $(0;\ b)$, то ее уравнение может быть записано в следующем виде:

$$\boxed{\frac{x}{a}+\frac{y}{b}=1}$$

(уравнение прямой в отрезках).

Прямые

$$Ax+By=C \qquad \text{и} \qquad A'x+B'y=C'$$

параллельны, если параллельны векторы нормалей к ним, то есть если существует такое число λ, что

$$A'=\lambda A \qquad \text{и} \qquad B'=\lambda B\,.$$

Прямые перпендикулярны друг другу, если векторы нормалей к ним взаимно ортогональны, то есть их скалярное произведение равно 0:

$$A\cdot A'+B\cdot B'=0 \qquad \text{или} \qquad \frac{A}{B}=-\frac{B'}{A'}\,.$$

Вектор $\overrightarrow{P_1P_2}$, параллельный прямой, которая проходит через точки $P_1(x_1;\ y_1)$ и $P_2(x_2;\ y_2)$, и называемый *направляющим вектором* прямой, имеет координаты $\overrightarrow{P_1P_2}(x_2-x_1;\ y_2-y_1)$. Ему перпендикулярен вектор $\mathbf{n}$, имеющий координаты $(y_2-y_1;\ -x_2-x_1)$, поэтому *уравнение прямой* P_1P_2 имеет вид:

$$\boxed{(y_2-y_1)x-(x_2-x_1)y=(y_2-y_1)x_1-(x_2-x_1)y_1\,.}$$

Последнее можно переписать также в виде:

$$\boxed{y=y_1=\frac{y_2-y_1}{x_2-x_1}(x-x_1)\,,}$$

или

$$\boxed{\begin{vmatrix} x & y & 1\\ x_1 & y_1 & 1\\ x_2 & y_2 & 1\end{vmatrix}=0.}$$

Если вектор нормали, присутствующий в уравнении прямой, является единичным, то, перенося постоянный член в левую часть, получаем так называемое *нормальное уравнение* прямой:

$$\boxed{A_0x + B_0y - C_0 = 0, \qquad \text{где} \qquad A_0^2 + B_0^2 = 1.}$$

Пусть $P_0(x_0;\ y_0)$ — произвольная точка плоскости. Расстояние d от точки P_0 до прямой определяется по формуле

$$\boxed{d = |A_0x_0 + B_0y_0 - C_0|.}$$

Пример: Чтобы найти расстояние от точки $P_0(-5;\ 10)$ до прямой $3x - 4y = 20$, разделим уравнение прямой на модуль вектора нормали $\mathbf{n}(3;\ -4)$. Таким образом мы получим нормальное уравнение прямой:

$$|\mathbf{n}| = \sqrt{3^2 + (-4)^2} = 5,$$

$$\frac{3}{5}x - \frac{4}{5}y - 4 = 0.$$

Подставляя в левую часть этого уравнения координаты точки P_0, получим искомое расстояние:

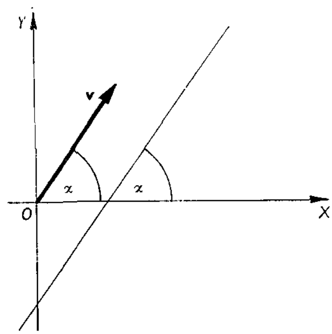
$$d = \left|\frac{3}{5}(-5) - \frac{4}{5}\cdot 10 - 4\right| = 15.$$

Если $B \neq 0$, то есть прямая не параллельна оси y, то, разделив уравнение прямой на B, можно привести его к виду:

$$y = -\frac{A}{B}x + \frac{C}{A}.$$

Число $-\dfrac{A}{B}$ мы обозначим через m. Геометрически оно выражает тангенс *направляющего угла* прямой (то есть угла между прямой и положительным направлением оси X) и называется *угловым коэффициентом прямой*. Число $\dfrac{C}{A}$ принято обозначать через b; прямая пересекает ось y в точке $(0;\ b)$. В этих обозначениях имеем:

$$\boxed{y = mx + b.}$$



Направляющий вектор **v** и направляющий угол α некоторой прямой

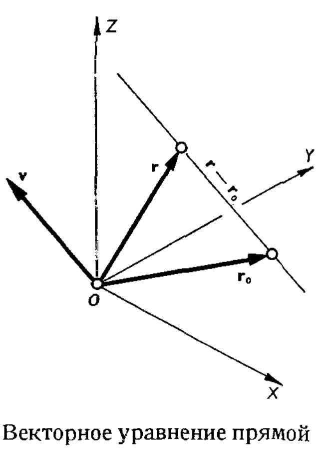
Если две прямые имеют угловые коэффициенты m_1 и m_2, то при $m_1 = m_2$ они параллельны, а при $m_1m_2 = -1$ взаимно перпендикулярны. Если прямые не перпендикулярны, то угол ω между ними определяется по формуле

$$\boxed{\operatorname{tg}\omega = \frac{m_2 - m_1}{1 + m_1m_2}.}$$

Чтобы определить точку пересечения двух прямых, нужно найти такую пару координат $(x;\ y)$, которая удовлетворяла бы обоим уравнениям, то есть необходимо решить систему двух уравнений первой степени с двумя неизвестными, которую образуют уравнения этих прямых. Все утверждения, касающиеся прямых и их пересечений, можно выразить с помощью уравнений прямых; тем самым всякую такую задачу можно свести к арифметической задаче. Таким изучением геометрических соотношений методами арифметики занимается *аналитическая геометрия*. Для нее характерно, что геометрические фигуры задаются уравнениями (или системами уравнений). Уравнением геометрической фигуры мы называем такое уравнение, которому удовлетворяют координаты точек этой фигуры и только они. Например, $y = x$ есть уравнение прямой, проходящей через начало координат, направляющий угол которой равен 45°. В то же время уравнение $y^2 = x^2$ уже не является уравнением этой прямой, так как ему удовлетворяют не только точки прямой $y = x$, но, например, и точка $(-1;\ 1)$.

ПАРАМЕТРИЧЕСКОЕ УРАВНЕНИЕ ПРЯМОЙ; УРАВНЕНИЕ ПЛОСКОСТИ

Если прямая рассматривается в пространственной системе координат, то для ее задания недостаточно одной точки и вектора нормали, так как в пространстве прямые, проходящие через данную точку и перпендикулярные данному вектору, заполняют целую плоскость. В то же время и в этом случае прямая однозначно определяется заданием некоторой ее точки и *направляющего вектора*, то есть вектора, параллельного прямой.



Векторное уравнение прямой

Пусть $\mathbf{r}_0(x_0;\ y_0;\ z_0)$ — радиус-вектор некоторой фиксированной точки P_0 прямой, а $\mathbf{v}(a;\ b;\ c)$ — направляющий вектор прямой. Если обозначить через $\mathbf{r}(x;\ y;\ z)$ радиус-вектор произвольной точки P прямой, то вектор $\mathbf{r}-\mathbf{r}_0$ параллелен прямой, а следовательно, и вектору $\mathbf{v}$, но тогда он представим в виде произведения вектора $\mathbf{v}$ на некоторое число, которое обозначим через t:

$$\boxed{\mathbf{r}-\mathbf{r}_0 = t\mathbf{v}.}$$

Число t называется *параметром* (вспомогательной переменной), а уравнение

$$\boxed{\mathbf{r} = \mathbf{r}_0 + t\mathbf{v}}$$

— *параметрическим уравнением* прямой в *векторной форме.* Если t принимает всевозможные вещественные значения, то конец вектора $\mathbf{r}$ пробегает всю прямую. Аналогичным уравнениям удовлетворяют и соответственные координаты этих векторов:

$$\boxed{\begin{aligned} x &= x_0 + at, \\ y &= y_0 + bt, \\ z &= z_0 + ct. \end{aligned}}$$

Это так называемая *система параметрических уравнений* прямой. Если ни одно из чисел a, b, c не равно 0 (то есть прямая не параллельна ни одной из координатных осей), то эту систему можно записать в следующем виде:

$$\boxed{\frac{x-x_0}{a} = \frac{y-y_0}{b} = \frac{z-z_0}{c}.}$$

Заметим, что прямую можно задать параметрическими уравнениями и в системе координат на плоскости. В этом случае, разумеется, имеем только два уравнения:

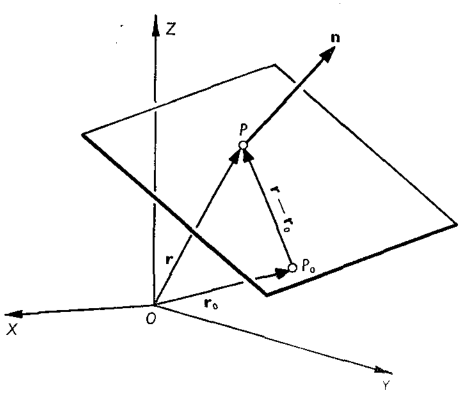
$$\boxed{\begin{aligned} x &= x_0 + at, \\ y &= y_0 + bt. \end{aligned}}$$

Исключая из этих уравнений t, получаем уже известное уравнение:

$$\boxed{b(x-x_0) - a(y-y_0) = 0.}$$

Вектор $(b;\ -a)$ нормали к прямой действительно ортогонален направляющему вектору $(a;\ b)$ прямой, так как их скалярное произведение равно 0.

Произвольную плоскость можно задать *вектором нормали* к ней и некоторой ее точкой. Действительно, если задать некоторую точку и некоторое направление, то в пространстве можно построить только одну плоскость, проходящую через данную точку перпендикулярно данному



Уравнение плоскости

направлению. Пусть $\mathbf{r}_0$ обозначает радиус-вектор некоторой фиксированной точки $P_0(x_0; y_0; z_0)$ плоскости, а $\mathbf{n}(A; B; C)$ — вектор нормали к ней. Обозначим через $\mathbf{r}$ радиус-вектор произвольной точки $P(x; y; z)$ плоскости. Поскольку вектор $\mathbf{r}-\mathbf{r}_0$ лежит в рассматриваемой плоскости, то он ортогонален вектору $\mathbf{n}$, а следовательно, их скалярное произведение равно 0:

$$\boxed{(\mathbf{r}-\mathbf{r}_0)\mathbf{n} = 0.}$$

Это так называемое *векторное уравнение* плоскости, ему удовлетворяют те и только те векторы $\mathbf{r}$, которые являются радиус-векторами точек плоскости. Переписывая это уравнение для координат векторов, получаем:

$$\boxed{A(x-x_0) + B(y-y_0) + C(z-z_0) = 0.}$$

Это уравнение принято записывать также в виде

$$\boxed{Ax + By + Cz = Ax_0 + By_0 + Cz_0}$$

или, если обозначить постоянную в правой части через D, то

$$\boxed{Ax + By + Cz = D.}$$

Если плоскость пересекает оси X, Y и Z, соответственно, в точках a, b и c, то ее уравнение можно записать в следующем виде:

$$\boxed{\frac{x}{a} + \frac{y}{b} + \frac{z}{c} = 1.}$$

Уравнение плоскости, проходящей через три точки

$$P_1(x_1; y_1; z_1), \quad P_2(x_2; y_2; z_2) \quad \text{и} \quad P_3(x_3; y_3; z_3)$$

может быть представлено так:

$$\begin{vmatrix} x & y & z & 1 \\ x_1 & y_1 & z_1 & 1 \\ x_2 & y_2 & z_2 & 1 \\ x_3 & y_3 & z_3 & 1 \end{vmatrix} = 0.$$

АНАЛИТИЧЕСКОЕ ВЫРАЖЕНИЕ ПЛОЩАДИ И ОБЪЕМА

При рассмотрении векторного и смешанного произведения векторов мы указали на их связь с вычислением площадей и объемов. На этом основываются и приводимые ниже соотношения. Площадь треугольника с вершинами в точках $P_1(x_1;\ y_1)$, $P_2(x_2;\ y_2)$ и $P_3(x_3;\ y_3)$ может быть найдена по следующей формуле:

$$t = \frac{1}{2}\begin{vmatrix} x_1 & y_1 & 1 \\ x_2 & y_2 & 1 \\ x_3 & y_3 & 1 \end{vmatrix}$$

Величина площади здесь получается со знаком плюс или минус, то есть, по сути говоря, следовало бы взять абсолютную величину определителя. Если $t = 0$, то точки P_1, P_2 и P_3 лежат на одной прямой.

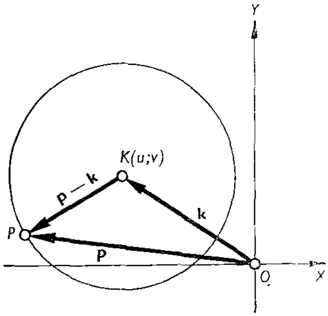
Объем тетраэдра с вершинами в точках $P_1(x_1;\ y_1;\ z_1)$, $P_2(x_2,\ y_2,\ z_2)$ $P_3(x_3,\ y_3,\ z_3)$ и $P_4(x_4,\ y_4,\ z_4)$ вычисляется по формуле

$$v = \frac{1}{6}\begin{vmatrix} x_1 & y_1 & z_1 & 1 \\ x_2 & y_2 & z_2 & 1 \\ x_3 & y_3 & z_3 & 1 \\ x_4 & y_4 & z_4 & 1 \end{vmatrix}$$

Величина этого объема также получается со знаком. Точки P_1, P_2, P_3 и P_4 тогда и только тогда принадлежат одной плоскости, когда $v = 0$.

УРАВНЕНИЕ ОКРУЖНОСТИ

Для точек окружности радиуса r с центром в точке $K(u;\ v)$ характерно то, что все они лежат на расстоянии r от точки K. Это значит, что если обозначить через $\mathbf{k}$ радиус-вектор точки K, а через $\mathbf{p}$ — радиус-вектор точки



Уравнение окружности

P, то для любой точки P окружности модуль вектора $\overrightarrow{PK}$ равен r, то есть

$$|\overrightarrow{KP}| = |\mathbf{p}-\mathbf{k}| = r.$$

Возводя это уравнение в квадрат, получим равносильное ему уравнение:

$$\boxed{(\mathbf{p}-\mathbf{k})^2 = r^2.}$$

Это *векторное уравнение* окружности. Учитывая, что вектор $\mathbf{p}-\mathbf{k}$ имеет координаты $(x-u;\ y-v)$, это уравнение можно переписать в координатной форме:

$$\boxed{(x-u)^2 + (y-v)^2 = r^2.}$$

Мы получили *уравнение окружности*. Если центр окружности совпадает с началом координат ($u = 0$, $v = 0$), то получаем так называемое *каноническое уравнение* окружности:

$$\boxed{x^2 + y^2 = r^2.}$$

В последнем случае легко записать уравнение касательной к окружности в некоторой ее точке $E(x_0;\ y_0)$. Радиус-вектор $\mathbf{e}$ точки E имеет координаты $(x_0;\ y_0)$ и перпендикулярен касательной, следовательно, уравнение касательной имеет вид:

$$x_0(x-x_0) + y_0(y-y_0) = 0.$$

После упрощений получаем:

$$\boxed{x_0x + y_0y = r^2.}$$

Уравнение вида

$$ax^2 + ay^2 + bx + cy + d = 0$$

является уравнением окружности, если его можно привести к виду:

$$(x-u)^2 + (y-v)^2 = r^2.$$

Это возможно при условии, что $a \neq 0$ и $b^2 + c^2 - 4ad > 0$. Действительно, в этом случае уравнение может быть приведено к виду:

$$\left(x+\frac{b}{2a}\right)^2+\left(y+\frac{c}{2a}\right)^2=\frac{b^2+c^2-4ad}{4a^2},$$

и, следовательно, задает окружность радиуса $\frac{b^2+c^2-4ad}{4a^2}$ с центром в точке $\left(-\frac{b}{2a};-\frac{c}{2a}\right)$.

УРАВНЕНИЯ ПЛОСКИХ СЕЧЕНИЙ КОНУСА

Если оси эллипса и гиперболы совпадают с осями координат, то эти кривые могут быть заданы так называемыми каноническими уравнениями.

Каноническое уравнение эллипса имеет вид:

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

где $2a$ и $2b$ — длины его осей.

Касательная к эллипсу в точке $(x_0; y_0)$ задается уравнением

$$\boxed{\frac{x_0}{a^2}x+\frac{y_0}{b^2}y=1.}$$

Если обозначить длину вещественной оси *гиперболы* через $2a$, расстояние между ее фокусами — через $2c$, а также ввести обозначение $b^2=c^2-a^2$, то каноническое уравнение примет следующий вид:

$$\boxed{\frac{x^2}{a^2}-\frac{y^2}{b^2}=1.}$$

Уравнение касательной к гиперболе в точке $(x_0; y_0)$ записывается так:

$$\boxed{\frac{x_0}{a^2}x-\frac{y_0}{b^2}y=1.}$$

Асимптоты гиперболы определяются уравнениями

$$\boxed{bx+ay=0 \quad \text{и} \quad bx-ay=0.}$$

При $a=b$ имеем равностороннюю гиперболу.

Параболу с параметром p расположим в системе координат так, чтобы ее ось совпадала с осью X, а вершина лежала в начале координат. В этом случае уравнение параболы имеет вид:

$$\boxed{y^2 = 2px.}$$

Касательная к ней в точке $(x_0;\ y_0)$ задается уравнением

$$\boxed{yy_0 = p(x + x_0).}$$

Для любого плоского сечения конуса половину длины хорды, проходящей через фокус перпендикулярно оси, на которой лежит фокус, называют параметром и обозначают через p. Для параболы параметр равен расстоянию от фокуса до директрисы. Если выбрать систему координат таким образом, чтобы ось X совпадала с осью сечения, проходящей через фокус, а начало координат — с одним из фокусов, то уравнение произвольного плоского сечения конуса можно записать в виде:

$$\boxed{x^2 + y^2 = (ex + p)^2,}$$

где различные сечения отличаются только значением e (эксцентриситетом). Если $0 < e < 1$, то имеем эллипс; при $e = 1$ получаем параболу; если же $e > 1$, то уравнение задает гиперболу (при $e = 0$ получается уравнение окружности).

КРИВЫЕ ВТОРОГО ПОРЯДКА

Как мы уже видели, плоские сечения конуса — включая и окружность — задаются уравнениями второй степени с двумя неизвестными.

Все линии, которые задаются уравнениями второй степени с двумя неизвестными, называются кривыми второго порядка.

Существуют и кривые второго порядка, не являющиеся плоскими сечениями конуса. Если, например, возвести в квадрат уравнение прямой

$$x + y = 0,$$

то мы получим уравнение

$$x^2 + y^2 + 2xy = 0.$$

Согласно данному определению, это уравнение кривой второго порядка, хотя эта линия состоит из тех же точек, что и прямая $x + y = 0$. Кривые второго порядка, содержащие прямую, называются вырожденными плоскими сечениями конуса. По сути говоря, уравнения вырожденных плоских сечений конуса получаются в результате перемножения уравнений двух прямых. Таким образом, вырожденные плоские сечения конуса представляют собой две пересекающиеся или две параллельные прямые, или же одну прямую, «взятую дважды». Однако любому уравнению второго порядка соответствует геометрически множество точек (быть может, и пустое), являющееся множеством общих точек конуса и некоторой плоскости.

Мы исключаем из рассмотрения кривые второго порядка, состоящие не более чем из одной точки. Например, кривая $x^2 + y^2 = 0$ состоит только из одной точки, а кривая $x^2 + y^2 = -1$ не имеет ни одной точки, так как сумма квадратов двух чисел не может быть отрицательной (в области вещественных чисел).

Общее уравнение кривой второго порядка имеет вид:

$$ax^2 + 2bxy + cy^2 + 2dx + 2ey + f = 0.$$

Определителем этой кривой называется определитель

$$A = \begin{vmatrix} a & b & d \\ b & c & e \\ d & e & f \end{vmatrix}$$

Обозначим его главный минор второго порядка через B:

$$B = \begin{vmatrix} a & b \\ b & c \end{vmatrix} = ac - b^2.$$

Кривая является вырожденной, если $A = 0$. Если при этом $B < 0$, то получаются две пересекающиеся прямые, если же $B = 0$ — то две параллельные или совпадающие прямые (в случае $B > 0$ кривая состоит максимум из одной точки).

Если $A \neq 0$, то общее уравнение задает уже известные нам плоские сечения конуса: при $B > 0$ — эллипс или окружность, при $B = 0$ — параболу, а при $B < 0$ — гиперболу.

Если кривая второго порядка не является вырожденной, то касательная к ней в точке $P_0(x_0; y_0)$ задается уравнением

$$(ax_0 + by_0 + d)x + (bx_0 + cy_0 + e)y + (dx_0 + ey_0 + f) = 0.$$

Координаты центра эллипса или гиперболы определяются системой уравнений

$$ax + by + d = 0.$$

$$bx + cy + e = 0.$$

Например: определителем кривой второго порядка, заданной уравнением

$$6x^2 + 7xy + 2y^2 - 3x - 2y - 10 = 0$$

является

$$A = \begin{vmatrix} 6 & 3{,}5 & 1{,}5 \\ 3{,}5 & 2 & -1 \\ -1{,}5 & -1 & -10 \end{vmatrix} \neq 0.$$

Следовательно, эта кривая не вырождена.

$$B = \begin{vmatrix} 6 & 3{,}5 \\ 3{,}5 & 2 \end{vmatrix} = -0{,}25 < 0.$$

Таким образом, кривая является гиперболой. Из системы уравнений

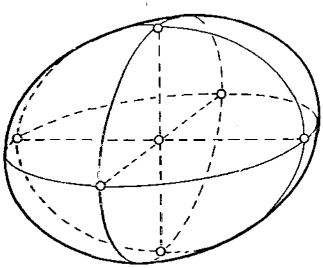
$$\begin{aligned} 6x + 3{,}5y - 1{,}5 &= 0 \\ 3{,}5x + 2y - 1 &= 0 \end{aligned}$$

находим координаты центра: $K(2;\ -3)$.

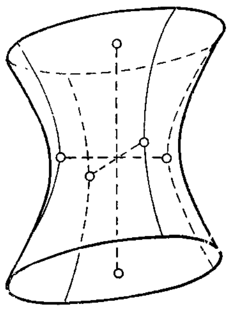
ПОВЕРХНОСТИ ВТОРОГО ПОРЯДКА

В пространстве кривым второго порядка соответствуют поверхности второго порядка. К ним относятся поверхности наиболее известных типов. Эти поверхности задаются уравнениями второй степени с тремя неизвестными.

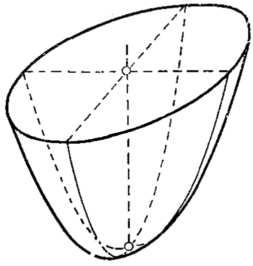
Поверхность второго порядка можно получить, например, в результате вращения некоторой кривой второго порядка (вырожденной или нет) вокруг ее оси симметрии. Это так называемые поверхности вращения. К ним относятся, например, *конус вращения*, *цилиндр вращения* и *сфера*.



Эллипсоид



Однополостный гиперболоид



Параболоид

Путем вращения эллипса вокруг одной из его осей получаем *элипсоид вращения*. Вращая гиперболу вокруг вещественной оси, получим *двуполостный гиперболоид вращения*. Если же вращать гиперболу вокруг мнимой оси (перпендикулярной вещественной), то получится *однополостный гиперболоид вращения*. При вращении параболы вокруг ее оси получается *параболоид вращения*.

Если каждую из этих поверхностей сжать перпендикулярно некоторой их плоскости симметрии, то мы получим, соотвественно, конус, цилиндр, эллипсоид, гиперболоид или параболоид второго порядка более общего вида.

Рассмотрим в пространстве три попарно скрещивающиеся прямые, параллельные данной плоскости. Всевозможные прямые, пересекающие все три рассмотренные прямые, образуют поверхность второго порядка, имеющую форму седла и называемую *гиперболическим параболоидом*.

Уравнения некоторых поверхностей второго порядка принимают простой вид, если, например, в качестве координатных плоскостей выбрать их плоскости симметрии. Приведем некоторые из таких уравнений. Сфера радиуса r с центром в начале координат задается уравнением

$$\boxed{x^2+y^2+z^2=r^2.}$$

Уравнение

$$\boxed{\frac{x^2}{a^2}+\frac{y^2}{b^2}+\frac{z^2}{c^2}=1}$$

задает эллипсоид. При $a=b$ имеем эллипсоид вращения.

Однополостный и двуполостный гиперболоиды задаются, соответственно, следующими уравнениями:

$$\boxed{\frac{x^2}{a^2}+\frac{y^2}{b^2}-\frac{z^2}{c^2}=1; \quad \frac{x^2}{a^2}-\frac{y^2}{b^2}-\frac{z^2}{c^2}=1.}$$

Уравнение эллиптического параболоида имеет вид:

$$\boxed{\frac{x^2}{a^2}+\frac{y^2}{b^2}=2z.}$$

Уравнение

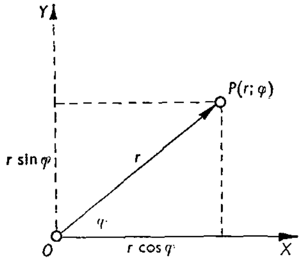
$$\boxed{\frac{x^2}{a^2}-\frac{y^2}{b^2}=2z \quad \text{или} \quad xy=z}$$

задает гиперболический параболоид.

ПОЛЯРНЫЕ КООРДИНАТЫ НА ПЛОСКОСТИ

Для описания некоторых фигур — главным образом тех, которые могут быть получены в результате вращения — вместо декартовых координат часто удобнее бывает использовать так называемые *полярные координаты*.

Построим на плоскости полупрямую (полярную ось) с начальной точкой O.



Полярные координаты на плоскости

Получение произвольной точки P плоскости однозначно определяется расстоянием $OP = r$ и углом φ между отрезком OP и полярной осью. Пара чисел $(r; \varphi)$ называются полярными координатами точки P.

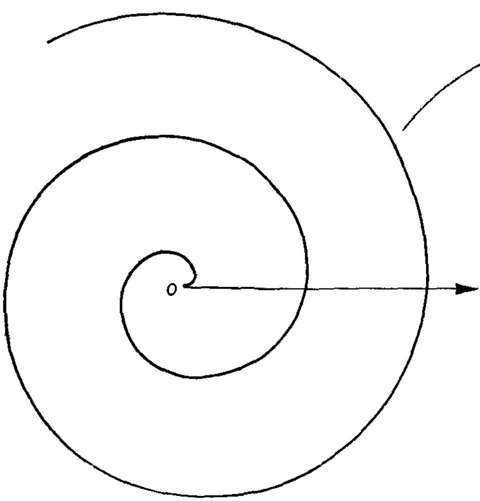
Если полярная ось совпадает с положительной полуосью OX декартовой системы координат, то между декартовыми координатами $(x; y)$ и полярными координатами $(r; \varphi)$ точки P имеют место следующие соотношения:

$$\boxed{x = r\cos\varphi; \qquad y = r\sin\varphi.}$$

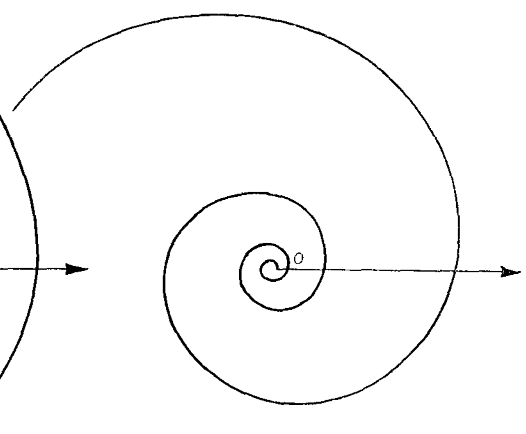
Окружность радиуса s, например, задается в полярных координатах уравнением

$$\boxed{r = s.}$$

Если для произвольного плоского сечения конуса выбрать точку O в фокусе, а за полярную ось принять ось сечения, проходящую через



Спираль Архимеда



Логарифмическая спираль

фокус, то уравнение сечения в полярных координатах будет иметь следующий вид:

$$r = \frac{p}{\pm 1 - e \cos \varphi}.$$

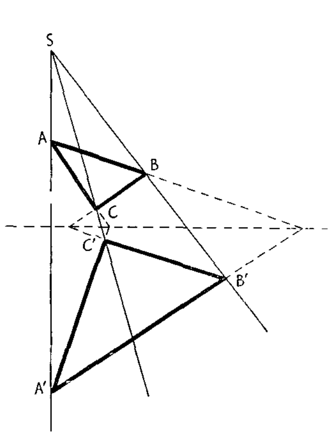
Здесь p — параметр плоского сечения конуса. Для эллипса $e < 1$, для параболы $e = 1$, а для гиперболы $e > 1$. В случае эллипса и параболы в знаменателе следует взять $+1$, а для двух ветвей гиперболы — соответственно, $+1$ и -1.

Особенно простой вид имеют в полярных координатах уравнения *спиралей*, задавать которые в другой системе координат существенно трудней. Архимедова спираль задается уравнением $r = a\varphi$, а логарифмическая спираль — уравнением $r = ae^{k\varphi}$. Для первой φ принимает неотрицательные, а для второй — любые вещественные значения. Логарифмическая спираль обладает одним интересным свойством: если вращать ее вокруг точки O, то создается впечатление, что спираль скручивается или раскручивается.

ПРОЕКТИВНАЯ ГЕОМЕТРИЯ

НЕСОБСТВЕННЫЕ ЭЛЕМЕНТЫ ПРОСТРАНСТВА; ТЕОРЕМА ДЕЗАРГА

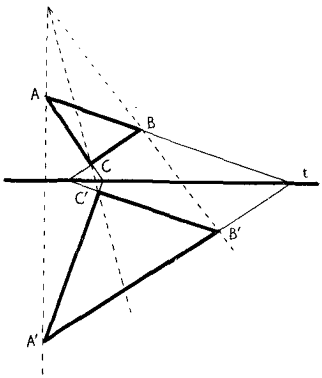
В процессе развития геометрии многие исследователи, теоретики и практики, — в первую очередь, художники ренессанса, исследовавшие законы перспективы, — неоднократно приходили к мысли, что параллельные прямые часто проявляют свойства пересекающихся прямых. Например, перспективой (центральной проекцией) двух параллельных прямых чаще всего оказываются две пересекающиеся прямые; человеческий глаз нередко воспринимает параллельные прямые как пересекающиеся. Еще более заметна эта связь в таких теоремах геометрии, как, например, теорема Дезарга. Прежде, чем перейти к ее формулировке, введем два понятия, связанных с перспективой треугольников.



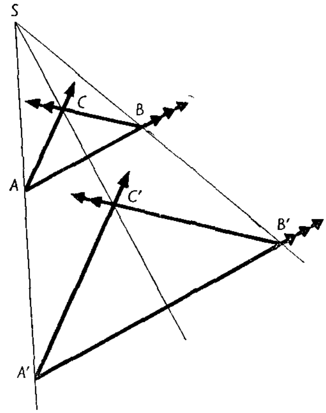
Треугольники, перспективные относительно точки

1. Назовем треугольники ABC и $A'B'C'$ перспективными относительно точки, если прямые AA', BB', CC' проходят через некоторую общую точку S, называемую центром.

2. Назовем треугольники ABC и $A'B'C'$ перспективными относительно оси, если пары прямых AB и $A'B'$, BC и $B'C'$, CA и $C'A'$ пересекаются на прямой t, называемой осью.

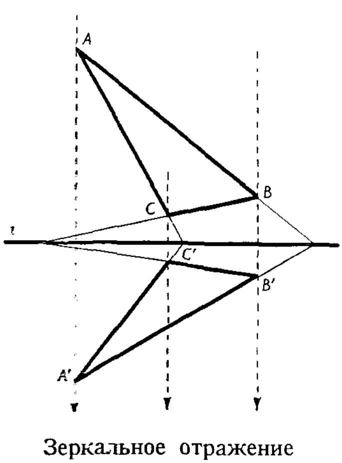


Треугольники, перспективные относительно оси



Перспективные подобные треугольники

Ж. Дезарг заметил, что в большинстве случаев справедливо следующее утверждение: два треугольника тогда и только тогда перспективны относительно некоторой точки, когда они перспективны и относительно некоторой оси. Однако в такой форме теорема верна не всегда. Например, некоторый треугольник и его образ при центрально-подобном преобразовании перспективны относительно точки S, являющейся центром подобия, но не могут быть перспективными относительно оси, так как соответственные стороны треугольников не пересекаются, а параллельны. Другой пример: произвольный треугольник и его зеркальный образ относительно некоторой прямой t являются перспективными относительно оси t, но не перспективны относительно точки, так как прямые, соединяющие соответственные вершины треугольников, параллельны. Нетрудно заметить, почему в этих примерах сформулированная выше теорема Дезарга оказывалась неверной: в обоих случаях препятствием послужила параллельность некоторых прямых. Чтобы избежать этого, стали говорить не о параллельных прямых, а о прямых, «пересекающихся в бесконечности». Используя это название, можно сказать, например, что прямые, соединяющие соответственные вершины треугольников, симметричных относительно оси, пересекаются в бесконечности, то есть треугольники являются перспективными относительно некоторой «бесконечно удаленной» точки.



Зеркальное отражение треугольника от оси

Следует подчеркнуть, что это только название, в обычном смысле параллельные прямые не имеют общей точки. Речь идет лишь о том, что вместо слова «параллельные» мы используем выражение «пересекаются в бесконечности».

Сформулируем сказанное несколько точнее. По существу такой подход означает, что всякую прямую мы дополняем некоторой воображаемой точкой, которую называют также бесконечно удаленной или *несобственной точкой*. В некотором смысле такое расширение можно сравнить с процессом расширения натурального ряда чисел, когда к нему добавляется 0. Число 0 уже не имеет такого «осязаемого» смысла, как числа 1, 2, 3, . . ., но тем не менее оказывается весьма полезным при вычислениях. (Еще больше общего это расширение имеет с дополнением множества вещественных чисел мнимыми числами.)

Отметим, что к каждой прямой добавляется одна несобственная точка, причем к параллельным прямым добавляется одна и та же несобственная точка. Таким образом несобственная точка может быть задана некоторым направлением, то есть любой из параллельных прямых, соответствующих этому направлению.

Совокупность несобственных точек плоскости называется *несобственной прямой*.

Таким образом, две параллельные прямые на плоскости пересекаются в некоторой точке несобственной прямой; всякая прямая на плоскости пересекает несобственную прямую в несобственной точке. Параллельным плоскостям ставится в соотвествие одна и та же несобственная прямая, то есть параллельные плоскости «пересекаются» по несобственной прямой. Отсюда следует, что несобственную прямую можно задать положением плоскости в пространстве. Наконец, говорят, что множество несобственных точек пространства образует *несобственную плоскость*.

При таком подходе структура пространства существенно меняется. Если обычные и несобственные элементы пространства рассматривать равноправно, то нетрудно убедиться, что теорема Дезарга справедлива во всех случаях.

ПРОЕКТИВНОЕ ПРОСТРАНСТВО. ДВОЙСТВЕННОСТЬ

Более общее представление о структуре пространства, дополненного несобственными элементами, позволяют составить следующие предложения:

1. (На плоскости) через две точки проходит ровно одна прямая.	1. (На плоскости) две прямые имеют ровно одну точку пересечения.
2. (В пространстве) через две точки проходит ровно одна прямая.	2. (В пространстве) две плоскости пересекаются по одной прямой.
3. Всякая плоскость имеет с произвольной прямой по крайней мере одну общую точку.	3. Через всякую точку и произвольную прямую проходит по крайней мере одна плоскость.

Таким образом, взаимное расположение прямых и плоскостей существенно упрощается.

Плоскость и пространтсво, дополненные несобственными элементами, называются, соответственно, проективной плоскостью и проективным пространством; исследованием их структуры занимается проективная геометрия.

Итак, в проективной геометрии обыкновенные и несобственные элементы пространства рассматриваются равноправно.

Кроме приведенных выше предложений 1, следует отметить следующие соотношения, имеющие место на плоскости:

Три точки, не лежащие на одной прямой, определяют три прямые, проходящие через них.

Три прямые, не пересекающиеся в одной точке, определяют три точки, являющиеся точками их пересечения.

Нетрудно заметить, что эти две пары предложений удовлетворяют так называемому *принципу двойственности на плоскости*, имеющему место в проективной геометрии. Согласно этому принципу, из всякой теоремы, касающейся точек, прямых и их взаимного расположения на плоскости, можно без труда получить новую теорему. Для этого достаточно изменить формулировку исходной теоремы таким образом, чтобы вместо точек речь шла о прямых, а вместо прямых — о точках. Соответственно этому, вместо точки пересечения двух прямых, например, следует говорить о прямой, проходящей через две точки, и наоборот.

Принцип двойственности имеет место и в пространстве. В этом случае точкам соответствуют плоскости, плоскостям — точки, а прямым — прямые. Примерами могут служить предложения, записанные выше под номерами 2 и 3.

ДВОЙНОЕ ОТНОШЕНИЕ

Это вспомогательное понятие, часто используемое в проективной геометрии.

Под двойным отношением четырех различных точек A, B, C и D некоторой прямой (обозначается: $(ABCD)$) понимают следующее отношение

$$(ABCD) = \frac{AC}{CB} : \frac{AD}{DB}.$$

Двойственным понятием для двойного отношения четырех точек прямой является двойное отношение четырех прямых, проходящих через одну точку. Двойное отношение (обозначается: $(abcd)$) прямых a, b, c и d, проходящих через одну точку, определяется следующим образом:

$$(abcd) = \frac{\sin(ac)}{\sin(cb)} : \frac{\sin(ad)}{\sin(db)}.$$

(Через (eg) мы обозначаем угол между прямыми e и g.) Двойное отношение точек и прямых зависит от их порядка; изменение порядка, вообще говоря, приводит к изменению двойного отношения. Понятие двойного отношения точек можно обобщить и на тот случай, когда одна из точек является несобственной. Если, например, D_i является несобственной точкой прямой, на котопой лежат точки A, B и C, то

$$\boxed{(ABCD_i) = \frac{AC}{CB}.}$$

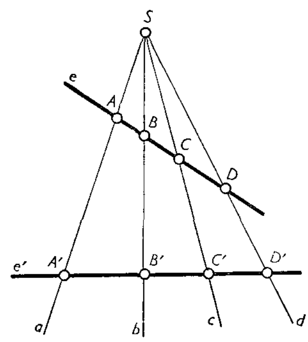
Если общая точка прямых a, b, c и d является несобственной, то их двойное отношение определяется следующим образом: пересечем эти прямые произвольной прямой и назовем двойным отношением рассматриваемых прямых двойное отношение четырех точек пересечения.

Важную роль играет следующая теорема: если некоторая прямая e пересекает прямые a, b, c и d, проходящие через некоторую точку S, соответственно, в точках A, B, C и D, то

$$\boxed{(abcd) = (ABCD).}$$

Если теперь пересечь эти четыре прямые некоторой другой прямой e', то для полученных точек пересечения A', B', C' и D' будет выполнено аналогичное соотношение: $(abcd) = (A'B'C'D')$. Таким образом,

$$\boxed{(ABCD) = (A'B'C'D').}$$



Связь между двойными отношениями четырех точек и четырех прямых

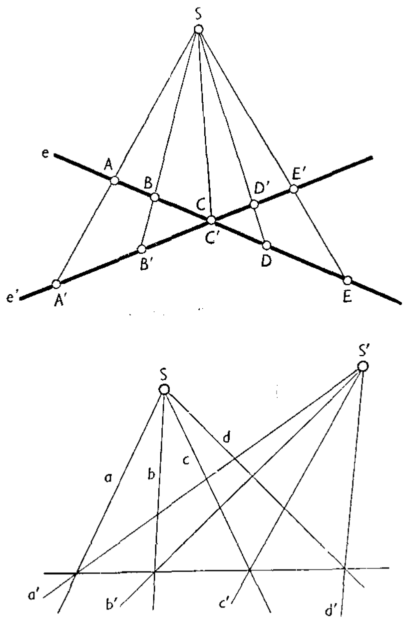
Поскольку точки A', B', C' и D' можно получить в результате центральной проекции точек A, B, C и D из точки S, то этот результат можно сформулировать следующим образом: при центральной проекции двойное отношение точек не меняется.

Если двойное отношение четырех точек или прямых равно -1, то такая четверка точек или прямых называется гармонической.

ПРОЕКТИВНОЕ ОТОБРАЖЕНИЕ

Проективным называется такое взаимно однозначное отображение (преобразование) проективной плоскости на проективную плоскость, при котором прямые переходят в прямые.

Проективное отображение осуществляют, например, центральные проекции одной плоскости на другую.



Перспективные ряды точек и пучки прямых

Проективное отображение однозначно определяется заданием четырех пар соответственных точек, никакие три из которых не лежат на одной прямой. Всякое проективное отображение сохраняет двойное отношение. Это означает, что двойное отношение четырех точек совпадает с двойным отношением их образов. Если делается различие между несобственными и обыкновенными элементами пространства, то проективные отображения включают в себя как частный случай аффинные преобразования, а следовательно, подобные преобразования и преобразования, сохраняющие расстояния.

Проективное отображение устанавливает между множеством точек произвольной прямой — назовем такое множество рядом точек — и рядом точек ее образа взаимно однозначное соответствие, сохраняющее двойное отношение. Два ряда точек, между которыми существует такое соответствие, мы будем называть проективными. Частным случаем проективных рядов точек являются так называемые перспективные ряды точек, то есть такие ряды, для которых прямые, соединяющие соответственные точки, пересекаются в одной точке, то есть один из рядов является проекцией другого.

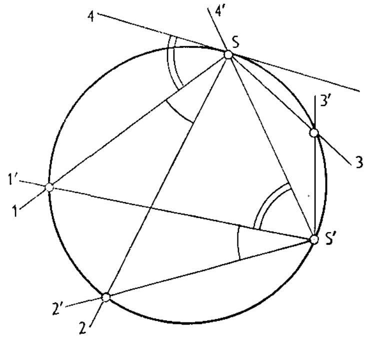
Для ряда точек двойственным понятием является пучок прямых, то есть множество прямых, проходящих через одну точку. Пучок прямых переходит при проективном отображении в пучок прямых; соответствие и в этом случае сохраняет двойное отношение. Два пучка прямых, между которыми существует такое соответствие, называются проективными. Частным случаем проективных пучков прямых являются так называемые перспективные пучки прямых, для которых все пары соответственных прямых пересекаются на одной прямой.

Проективное соответствие между двумя рядами точек или двумя пучками прямых однозначно определяется заданием трех пар соответственных элементов.

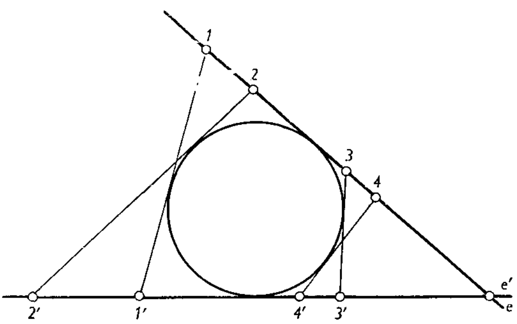
ПОЛУЧЕНИЕ ПЛОСКИХ СЕЧЕНИЙ КОНУСА ПРИ ПОМОЩИ ПРОЕКТИВНЫХ ПРЕОБРАЗОВАНИЙ

Проективные пучки прямых или ряды точек нетрудно построить с помощью окружности. Выберем на окружности две точки S и S' и соединим каждую из них прямыми со всеми точками окружности. Мы получим два пучка прямых, между элементами которых существует взаимно однозначное соотвествие: соответственными являются прямые, пересекающиеся на окружности. Общим элементом двух пучков является прямая SS'; ей соответствует касательная к окружности. Указанное соответствие сохраняет двойное отношение. Действительно, из свойств вписанных углов следует, что угол между двумя произвольными прямыми одного пучка равен углу между соответствующими им прямыми другого пучка, значит, равны и синусы этих углов, а отсюда сразу следует, что соответствие сохраняет двойное отношение. Таким образом рассматриваемые пучки прямых являются проективными. Отсюда вытекает также, что окружность может быть получена как множество точек пересечения соответственных элементов двух проективных пучков прямых. Поскольку всякое плоское сечение конуса может быть получено как некоторая проекция окружности, то можно показать, что всякое плоское сечение конуса может быть получено при помощи двух проективных пучков прямых. Обратно: точки пересечения соответственных элементов двух проективных (но не перспективных) пучков прямых образуют некоторое плоское сечение конуса.

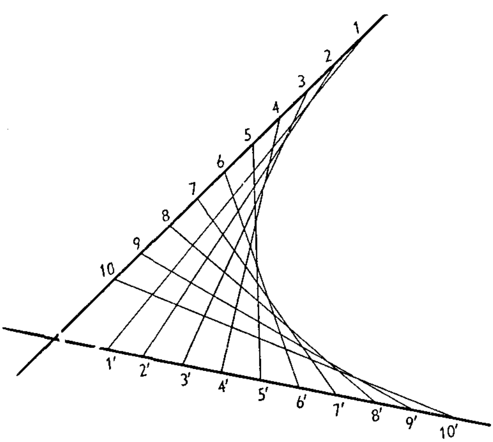
Пусть теперь e и e' — две произвольные касательные к данной окружности. Тогда все остальные касательные устанавливают взаимно однозначное соотвествие между рядами точек e и e'. Соответственными являются точки пересечения прямых e и e' с одной и той же касательной. Аналогич-



Окружность и проективные пучки прямых



Окружность и проективные ряды точек



Касательные к параболе

но предыдущему, можно показать, что ряды точек e и e' являются проективными, таким образом, множество всех касательных к данной окружности представляет собой совокупность прямых, соединяющих соответственные точки двух проективных рядов точек. Семейство касательных к произвольному плоскому сечению конуса может быть представлено как совокупность прямых, соединяющих соотвественные точки двух проективных рядов точек. Обратно: совокупность прямых, соединяющих соответственные точки двух проективных рядов точек, представляет собой семейство касательных к некоторому плоскому сечению конуса.

Если два ряда точек подобны, то есть равны отношения их соответственных отрезков, то эти ряды проективны; в частности, проективными являются два совместимых (равных) ряда точек, причем если эти ряды не перспективны, то прямые, соединяющие их соответственные точки, являются касательными к некоторой параболе.

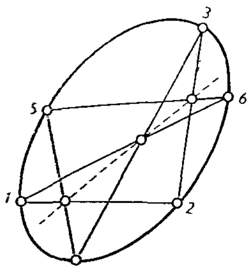
ТЕОРЕМЫ ПАСКАЛЯ И БРИАНШОНА

Эти две теоремы, пожалуй, лучше всего раскрывают структуру плоских сечений конуса, устанавливая общие свойства, характерные для всех этих кривых.

Шестиугольник называется *шестиугольником Паскаля*, если точки пересечения прямых, являющихся продолжениями его противоположных сторон, лежат на одной прямой. Если пронумеровать вершины шестиугольника числами 1, 2, 3, 4, 5, 6, то противоположными считаются следующие пары сторон: (12, 45), (23, 56), (34, 61).

Теорема Паскаля. Для того, чтобы некоторый шестиугольник был шестиугольником Паскаля, необходимо и достаточно, чтобы его вершины принадлежали некоторому плоскому сечению конуса. (Предполагается, что никакие три вершины шестиугольника не лежат на одной прямой).

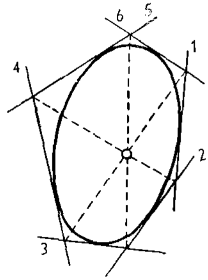
Эта теорема дает возможность по пяти данным точкам построить остальные точки плоского сечения конуса. Действительно, если рассмотреть шестиугольник Паскаля, пять вершин которого совпадают с данными точками, то и шестая вершина должна лежать на искомой кривой. С другой стороны, если зафиксировать пять вершин шестиугольника Паскаля, то шестую вершину можно выбрать бесконечно многими способами, она и опишет всю кривую. Заметим, что теорема Паскаля остается верной и в том случае, когда некоторые из вершин шестиугольника совпадают. В этом случае в качестве прямой, соединяющей две совпадающие (последовательно пронумерованные) вершины, следует брать касательную к кривой в этой точке.



Теорема Паскаля

Шестиугольник называется *шестиугольником*

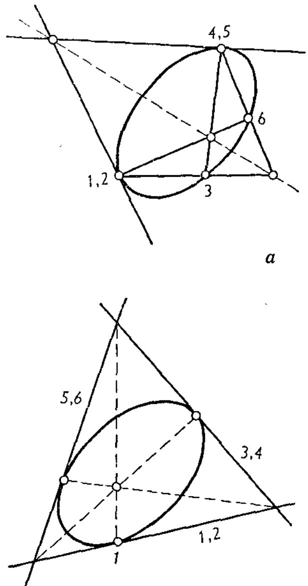
Брианшона, если при некоторой нумерации его сторон точки пересечения прямых, соединяющих противоположные вершины, лежат на одной прямой. Аналогично предыдущему, противоположными считаются точки пересечения следующих пар сторон: (12, 45), (23, 56) и (34, 61).



Теорема Брианшона

Теорема Брианшона. Для того, чтобы шестиугольник был шестиугольником Брианшона, необходимо и достаточно, чтобы его стороны (их продолжения) были касательными к некоторому плоскому сечению конуса. (Предполагается, что никакие три из этих прямых не пересекаются в одной точке).

При помощи теоремы Брианшона на основе пяти данных касательных к некоторому плоскому сечению конуса можно построить сколь угодно много касательных. Если даны пять сторон шестиугольника Брианшона, то шестую сторону можно построить бесконечно многими способами. Теорема остается верной и в том случае, если некоторые из сторон шестиугольника совпадают. В этом случае точкой пересечения совпадающих касательных следует считать точку касания.



Частные случаи треугольников Паскаля (*a*) и Брианшона (*b*)

Из сказанного выше следует, что всякое плоское сечение конуса однозначно определяется следующими данными:

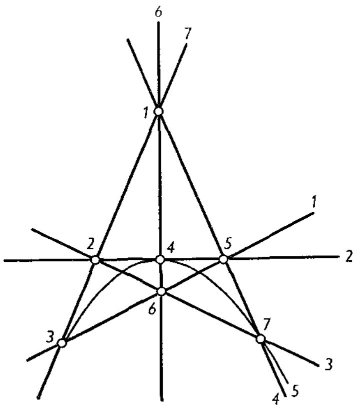
1. пятью точками;
2. четырьмя точками и касательной в одной из них;
3. тремя точками и касательными в двух из них;
4. пятью касательными;
5. четырьмя касательными и одной из точек касания;
6. тремя касательными и двумя из точек касания.

КОНЕЧНЫЕ ПРОЕКТИВНЫЕ ПЛОСКОСТИ

Проективная геометрия возникла из евклидовой и долгое время рассматривалась не как самостоятельная геометрия, а как новая трактовка евклидовой геометрии. Однако если построить проективную геометрию независимо от евклидовой, основываясь на другой аксиоматике, то различие между ними становится очевидным. В случае плоскости наиболее существенное различие состоит в том, что, в отличие от евклидовой геометрии, на проективной плоскости любые две прямые пересекаются.

При аксиоматическом построении проективной геометрии проективной плоскостью называют множество точек и прямых, которые удовлетворяют следующим аксиомам:

1. Через две различные точки можно провести одну и только одну прямую;
2. Две различные прямые имеют одну и только одну общую точку.
3. Во множестве существуют четыре точки, никакие три из которых не лежат на одной прямой.



Модель проективной плоскости, состоящей из семи точек

Разумеется, из этих аксиом можно вывести лишь очень немного теорем, поэтому принято вводить еще дополнительные аксиомы. Нетрудно заметить, что аксиомы, определяющие проективную плоскость, вовсе не требуют, чтобы прямая содержала бесконечно много точек. Действительно, можно указать примеры конечных множеств, элементы которых называют точками или прямыми, и для которых определено понятие «принадлежности», причем эти точки и прямые удовлетворяют аксиомам проективной плоскости. Такие конечные множества называют *конечными проективными плоскостями*.

Прямые в такой «искусственной геометрии» нельзя изобразить прямыми на обыкновенной плоскости.

Простейшая из таких конечных плоскостей состоит из 7 точек и 7 прямых. На рисунке видно, что через любые две точки проходит ровно одна прямая, и любые две прямые имеют ровно одну точку пересечения.

Для наглядности такую плоскость можно задать с помощью специальной таблицы. Пусть строки квадратной таблицы обозначают прямые $e_1, e_2, \ldots, e_n$, а столбцы — точки $P_1, P_2 \ldots, P_n$ плоскости.

Если точка принадлежит прямой, то на пересечении соотвествующей строки и столбца запишем знак $\times$, а остальные клетки таблицы оставим пустыми.

Нетрудно убедиться, что приводимая ниже таблица задает плоскость, через любые две точки которой проходит ровно одна прямая (например, точки P_4 и P_9 соединяет прямая e_{13}) и любые две прямые которой пересекаются ровно в одной точке (например, прямые e_5 и e_8 пересекаются в точке P_5).

	P_1	P_2	P_3	P_4	P_5	P_6	P_7	P_8	P_9	P_{10}	P_{11}	P_{12}	P_{13}
e_1	×	×	×	×									
e_2	×				×	×	×						
e_3	×							×	×	×			
e_4	×										×	×	×
e_5		×			×			×			×		
e_6		×				×			×			×	
e_7		×					×			×			×
e_8			×		×				×				×
e_9			×			×				×	×		
e_{10}			×				×	×				×	
e_{11}				×	×					×		×	
e_{12}				×		×		×					×
e_{13}				×			×		×		×		

Для всех конечных проективных плоскостей характерна своеобразная «правильность» такой таблицы. Если некоторой прямой принадлежит k точек, то каждой прямой на этой плоскости принадлежит k точек и через каждую точку проходит k прямых. Число n точек равно числу прямых, причем $n = k^2 - k + 1$. Приведенная таблица соотвествует случаю $k = 4$, $n = 13$. До сегодняшнего дня остается невыясненным, для каких значений k можно построить такую таблицу (то есть плоскость). Предполагают, что такое построение осуществимо, если число $k-1$ является простым или некоторой целой степенью простого числа. Доказано, например, что для $k = 7$ такую плоскость построить нельзя.

В заключение заметим, что исследованием конечных проективных плоскостей занимаются геометрия и комбинаторика.

НЕЕВКЛИДОВЫ ГЕОМЕТРИИ

ПРОБЛЕМА ПАРАЛЛЕЛЬНОСТИ

На евклидовой плоскости две прямые называются параллельными, если они не имеют общей точки. Основное свойство параллельности в евклидовом смысле зафиксировано в аксиоме V Евклида (в приведенном нами перечне (см. стр. 242) она также присутствовала под номером V). По существу эта аксиома утверждает, что на плоскости через данную точку можно провести только одну прямую, не пересекающую данную прямую. Эта аксиома явилась отправной точкой многих исследований, главным образом потому, что ее утверждение выходит за пределы человеческого восприятия и опыта. Таким образом, закономерной была бы мысль, что если одну из двух прямых, считающихся параллельными, повернуть вокруг некоторой ее точки на очень малый угол, то может случиться, что и эта сдвинутая прямая не будет пересекать вторую прямую ; другими словами, может оказаться, что через данную точку можно провести более одной прямой, не пересекающей данную прямую. Вначале справедливость этой спорной аксиомы пытались проверить, основываясь на опыте. С этой целью проводились исследования таких следствий аксиомы параллельности, критерием справедливости которых может служить человеческий опыт. Этой аксиоме равносильна, например, теорема, утверждающая, что сумма (внутренних) углов треугольника равна 180°. Это значит, что если выполняется аксиома V, то сумма углов треугольника равна 180° ; обратно : если сумма углов произвольного треугольника равна 180°, то должна выполняться аксиома V.

В процессе этих исследований Лежандр доказал, что если из аксиом Евклида исключить аксиому V, то на основе остальных аксиом (так называемой *остаточной системы аксиом)* можно показать, что сумма углов треугольника не может превышать 180°. Больше того, ему удалось доказать, что если сумма углов некоторого треугольника меньше 180°, то это верно и для любого треугольника ; если же для некоторого треугольника сумма углов равна 180°, то и сумма углов любого треугольника равна 180°. Таким образом всю проблему можно было бы решить, точно измерив сумму углов одного-единственного треугольника. Однако не может существовать абсолютно точных измерительных приборов, что не позволяет решить эту проблему путем измерений.

Отметим, что теоремы, которые могут быть доказаны на основе остаточной системы аксиом, называют *абсолютными*. Абсолютными теоремами являются, например, признаки равенства треугольников, а также теоремы, утверждающие, что медианы треугольника пересекаются в одной точке, что внешний угол треугольника больше любого внутреннего угла, не смежного с ним, что из данной точки на данную прямую можно опустить один перпендикуляр и другие.

ГИПЕРБОЛИЧЕСКАЯ ГЕОМЕТРИЯ БОЙАИ—ЛОБАЧЕВСКОГО

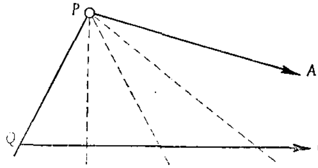
В 20-х годах прошлого века венгерский математик Янош Бойаи и русский математик Н. Е. Лобачевский независимо друг от друга пришли к заключению, что геометрию можно построить и основываясь на системе аксиом, в которую вместо аксиомы V включена противоположная ей аксиома. Эта аксиома утверждает следующее : в плоскости, определяемой некоторой прямой a и точкой P, не лежащей на этой прямой, через точку P можно провести более одной прямой, не пересекающей прямую a.

Такая геометрия получила название *геометрии Бойаи—Лобачевского*, или *гиперболической геометрии*. Позднее были разработаны и другие неевклидовы геометрические системы, например, геометрия Римана, однако, здесь мы не станем на них останавливаться.

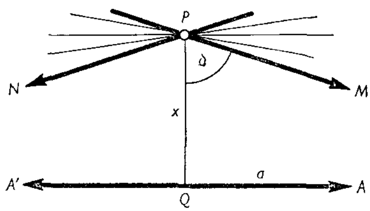
В гиперболической геометрии также существует понятие параллельности, определение которого дается ниже. Заметим прежде всего, что если P — начальная точка некоторого луча, а A — некоторая произвольная точка того же луча, то мы будем говорить о луче PA. Пусть теперь PA и QB — два луча, обладающие свойствами:

1. PA и QB расположены по одну сторону от прямой PQ;
2. PA и QB не пересекаются;
3. всякая прямая, проходящая через точку P в угловой области QPA, пересекает луч QB.

В этом случае говорят, что PA и QB параллельны.



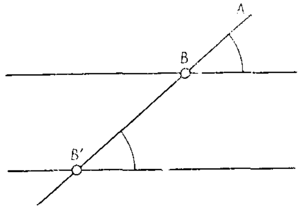
Параллельные прямые на гиперболической плоскости



Параллельные и ультрапараллельные прямые

Такое определение параллельности является *абсолютным*, оно включает в себя и понятие параллельности в евклидовом смысле. Наглядно понятие параллельности означает, что если поворачивать луч PA вокруг точки P в сторону точки B, то вращающийся луч окажется параллельным QB как раз «в момент отрыва», то есть при первом положении, когда он не пересекает QB. Если PA и QB параллельны, то это обозначается так: $PA \parallel QB$.

Пусть теперь a — некоторая прямая и P — точка, *не лежащая* на этой прямой. Опустим из точки P перпендикуляр на прямую a; его основание Q делит прямую a на две полупрямых: QA и QA'. Если теперь из точки P провести лучи, соотвественно параллельные полупрямым QA и QA', то есть $PM \parallel QA$ и $PN \parallel QA'$, то эти лучи не могут лежать на одной прямой, так как тогда существовало бы только одна прямая, проходящая через точку P и не пересекающая прямую a. Таким образом, лучи PM и PN расположены симметрично относительно перпендикуляра PQ и вместе со своими продолжениями образуют две такие угловые области, в которых через точку P проходит бесконечно много прямых, не пересекающих прямую a; говорят, что эти прямые *ультрапараллельны* прямой a. Нетрудно заметить, что в гиперболической геометрии определяется параллельность полупрямых; две прямые считаются параллельными, если у них существуют параллельные полупрямые. Итак, через точку P можно провести две прямых, параллельных a, и бесконечно много ультрапараллельных прямых.



Ультрапараллельные прямые

Угол $\sphericalangle QPM = \delta$ называется углом параллельности, соответствующим расстоянию PQ. Этот угол всегда острый (в евклидовой геометрии — прямой). Можно показать, что параллельность не зависит от выбора начальной точки лучей, а также обладает свойствами симметрии (то есть если $PM \parallel QA$, то $QA \parallel PM$) и транзитивности (то есть если $PM \parallel QA$ и $QA \parallel RB$, то $PM \parallel RB$).

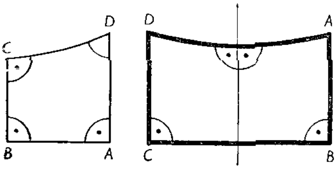
Если углы ABC и $AB'C'$ равны, и точки ABB' лежат на одной прямой, а точки C и C' лежат по одну сторону от прямой AB, то прямые BC и $B'C'$ ультрапараллельны. Это верно и в том случае, если каждая из двух прямых перпендикулярна некоторой третьей прямой. Заметим, что для двух ультрапараллельных прямых всегда существует прямая, перпендикулярная к каждой из них.

ЧЕТЫРЕХУГОЛЬНИКИ ЛАМБЕРТА И САККЕРИ

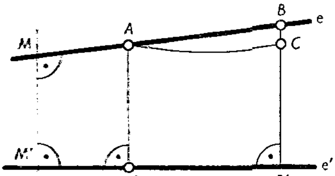
Сумма внутренних углов произвольного треугольника меньше 180°. Действительно, согласно теоремам Лежандра, сумма углов произвольного треугольника не может превышать 180°; однако, если бы она равнялась 180°, то выполнялась бы аксиома V. Отсюда следует также, что сумма углов четырехугольника меньше 360°.

Четырехугольник $ABCD$ называется *четырехугольником Ламберта*, если $\sphericalangle A = \sphericalangle B = \sphericalangle C = 90°$. Таким образом в этом случае $\sphericalangle D$ должен быть острым. Противоположные стороны такого четырехугольника ультрапараллельны, так как для них существуют общие перпендикуляры.

Четырехугольник $ABCD$ называется *четырехугольником Саккери*, если $AB = CD$ и стороны AB и CD перпендикулярны к BC. Прямая, перпендикулярная к стороне BC в ее середине, является осью симметрии четырехугольника, то есть является общим перпендикуляром для сторон BC и AD. Ось симметрии делит этот четырехугольник на два четырехугольника Ламберта, таким образом $\sphericalangle A = \sphericalangle D$ — острые углы, а стороны BC и AD ультрапараллельны. При удалении от общего перпендикуляра расстояние между ультрапараллельными прямыми увеличивается. В самом деле, пусть прямые e и e' ультрапараллельны, и пусть MM' — общий перпендикуляр к ним. Удаляясь от точки M вдоль прямой e, отметим точки A и B, и опустим из них перпендикуляры на прямую e'; основания перпендикуляров обозначим, соответственно, через A' и B'. Отложив расстояние AA' на $B'B$ от точки B', получим точку C. В четырехугольнике Ламберта $A'M'MA$ угол $\sphericalangle MAA'$ острый, то есть смежный с ним $\sphericalangle A'AB$ — тупой. В четырехугольнике Саккери $A'ACB'$ угол $\sphericalangle A'AC$ острый, поэтому прямая AC проходит внутри угловой области $A'AB$, то есть точка C расположена ближе к точке B', чем точка B, а следовательно, $BB' > CB' = AA'$.



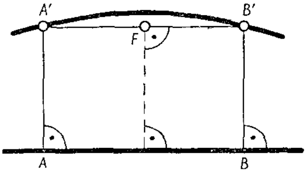
Четырехугольники Ламберта и Саккери



Расстояние между ультрапараллельными прямыми увеличивается

ЛИНИИ РАССТОЯНИЙ

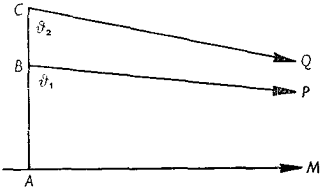
В евклидовой геометрии точки, расположенные по одну сторону от данной прямой на одинаковом расстоянии от нее, лежат на некоторой прямой, параллельной данной прямой. Нетрудно убедиться, что на гиперболической плоскости совокупность точек, расположенных по одну сторону от данной прямой на одинаковом расстоянии от нее (так называемая *линия расстояний* или *гиперцикл*), не является прямой. Действительно, выберем на прямой e произвольные точки A и B, восстановим из них перпендикуляры к e и отложим от их оснований равные отрезки, концы которых обозначим, соответственно, через A' и B'. Если бы линия расстояний была прямой, то она совпадала бы с прямой $A'B'$. Однако четырехугольник $AA'B'B$ является четырехугольником Саккери, поэтому прямые $A'B'$ и AB ультрапараллельны, а значит, середина F отрезка $A'B'$ расположена ближе к прямой e, чем точки A' и B'. Следовательно, линия расстояний не может совпадать с прямой $A'B'$. Таким образом, на гиперболической плоскости линия расстояний есть некоторая кривая.



Линия расстояний

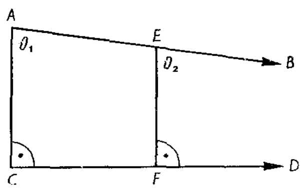
НЕКОТОРЫЕ СВОЙСТВА ПАРАЛЛЕЛЬНЫХ ПРЯМЫХ

Чем больше расстояние между параллельными прямыми, тем меньше соответствующий им угол параллельности. В самом деле, восстановим из точки A перпендикуляр к полупрямой AM, из двух точек B и C которого проведем полупрямые BP и CQ, параллельные AM. В силу транзитивности имеем: $CQ \parallel BP$. Пусть $\sphericalangle ABP = \vartheta_1$ и $\sphericalangle ACQ = \vartheta_2$. Тогда равенство $\vartheta_1 = \vartheta_2$ не может иметь места, так как в противном случае полупрямые CQ и BP были бы ультрапараллельны. Не может выполняться и неравенство $\vartheta_2 > \vartheta_1$, так как тогда полупрямую CQ можно было бы в результате поворота привести в некоторое положение CQ', при котором $\sphericalangle ACQ' = \vartheta_1$, но после поворота эта полупрямая должна была бы пересекать BP, что невозможно, поскольку из-за условия $\sphericalangle ACQ' = \vartheta_1$ полупрямые CQ' и BP были бы ультрапараллельны.



Чем больше расстояние между параллельными прямыми, тем меньше соответствующий угол параллельности

При движении вдоль параллельных прямых расстояние между ними уменьшается. Пусть $AB \parallel CD$ и $AC \perp CD$. Выберем на полупрямой AB точку E, которая расположена в направлении параллельности от точки A (см.



Расстояние между параллельными прямыми уменьшается рисунок). Пусть F — основание перпендикуляра, опущенного из точки E на CD. Обозначим $\sphericalangle CAB$ через ϑ_1, а $\sphericalangle FEB$ через ϑ_2. Углам ϑ_1 и ϑ_2 соответствуют расстояния AC и EF. Сумма углов четырехугольника $AEFC$ меньше 360°, то есть

$$90° + 90° + \vartheta_1 + (180° - \vartheta_2) < 360°,$$

откуда

$$\vartheta_1 < \vartheta_2.$$

Однако, согласно предыдущему, чем меньше угол параллельности, тем больше расстояние между параллельными прямыми, то есть $AC > EF$, таким образом, расстояние между полупрямыми действительно уменьшается.

МЕТРИЧЕСКИЕ СООТНОШЕНИЯ НА ГИПЕРБОЛИЧЕСКОЙ ПЛОСКОСТИ

Зависимость между углом параллельности и расстоянием между параллельными прямыми можно выразить и аналитически. Из системы аксиом гиперболической геометрии можно вывести, что расстояние x между параллельными прямыми и угол параллельности ϑ должны удовлетворять следующему соотношению:

$$\boxed{\operatorname{ctg}\frac{\vartheta}{2} = e^{\frac{x}{k}},}$$

где e — основание натурального логарифма ($e = 2{,}7182\ldots$) (см. «Математический анализ», стр. 459.), а k — некоторая постоянная характеристика геометрии; при различных значениях k получаются различные гиперболические геометрии. Роль этой постоянной похожа на роль радиуса сферы в сферической геометрии: при изменении радиуса всякий раз получаются разные сферические геометрии. По аналогии значение постоянной k принято называть также «кривизной» гиперболической плоскости. Постоянная k присутствует почти в каждом метрическом соотношении гиперболической геометрии.

Длину окружности радиуса r принято обозначать через $\bigcirc r$ (читается: окружность r); она определяется по формуле

$$\boxed{\bigcirc r = 2k\pi\, sh\frac{r}{k}.}$$

Гиперболические функции sh и ch(см. «Математический анализ», стр 401) очень часто присутствуют в формулах гиперболической геометрии, отсюда происходит и их название.

Вывод тригонометрических соотношений здесь вызывает большие трудности, нежели в евклидовой геометрии. Причина заключается в том, что в гиперболической геометрии отсутствует понятие подобия; если, например, два треугольника имеют соответственно равные углы, то они равны. Если гиперболическая геометрия строится аксиоматически, то тригонометрические функции здесь следует определять методами математического анализа (например, при помощи степенных рядов). Примечателен следующий результат, называемый *абсолютной теоремой синусов Бойаи*: для любого треугольника отношение длин окружностей, построенных на его сторонах, равно отношению противоположных им углов треугольника, то есть

$$\boxed{\bigcirc a : \bigcirc b = \sin\alpha : \sin\beta.}$$

(Это соотношение имеет место и в евклидовой геометрии, с той лишь разницей, что там длина окружности радиуса a равна $2a\pi$).

Если катеты прямоугольного треугольника равны a и b, а гипотенуза c, то

$$\boxed{ch\frac{a}{k}\cdot ch\frac{b}{k} = ch\frac{c}{k}\,.}$$

Это один из аналогов теоремы Пифагора на гиперболической плоскости.

Аналогия между гиперболической и сферической геометриями наблюдается и при вычислении площадей. Если α, β и γ обозначают углы некоторого треугольника, то его площадь

$$\boxed{t = k^2\,\delta,}$$

где $\delta = \pi - (\alpha + \beta + \gamma)$ — так называемый *дефект*. Площадь круга радиуса r (обозначается через $\odot\, r$) вычисляется по формуле

$$\boxed{\odot r = 4\pi k^2 sh^2\frac{r}{2k}\,.}$$

ГЕОМЕТРИЯ БОЙАИ—ЛОБАЧЕВСКОГО И ДЕЙСТВИТЕЛЬНОСТЬ

Первые создатели гиперболической геометрии исходили из того, что непосредственный опыт не противоречит предположению, что через данную точку можно провести более одной прямой, не пересекающей данную прямую. Такой подход приводит к выводу, что гиперболическая геометрия также может быть использована для описания окружающего нас пространства. Этому, казалось бы, противоречат теоремы и метрические соотношения, с которыми мы только что познакомились. Однако более глубокое исследование довольно скоро убеждает нас в том, что в принципе геометрия Бойаи—Лобачевского точно так же пригодна для описания материального мира, как и евклидова геометрия.

Непривычна, например, теорема, утверждающая, что множество точек, расположенных на одинаковом расстоянии от данной прямой по одну сторону от нее, лежат не на прямой, а на некоторой кривой линии. Однако это не противоречит опыту, согласно которому мы можем утверждать лишь, что, *если расстояния не слишком велики*, то точки, лежащие на одинаковом растоянии от данной прямой, принадлежат некоторой линии, не очень отличающиейся от прямой; последнее, в свою очередь, не противоречит теореме о расстоянии между параллельными прямыми.

Другим возражением может явиться тот факт, что метрические соотношения на гиперболической плоскости существенно отличаются от евклидовых, справедливость которых доказана опытом. Приведем пример, проще всего показывающий, что и здесь нет никакого противоречия.

Рассмотрим прямоугольный треугольник, катеты которого равны, соответственно 3 и 4 единицам измерения. Используя евклидову геометрию, по теореме Пифагора получаем, что гипотенуза равна 5 единицам измерения. Если мы хотим проверить этот результат непосредственными измерениями, то можем утверждать лишь, что длина гипотенузы приближенно равна 5 единицам измерения; из-за неточности измерительных приборов мы не располагаем абсолютно точными результатами измерений. Воспользуемся для вычисления длины гипотенузы соотношением $ch\frac{a}{k}\cdot ch\frac{b}{k}=ch\frac{c}{k}$, имеющим место на гиперболической плоскости.

Если постоянную характеристику k геометрии выбрать, например, равной 100, то для c получается значение 5,09; при $k = 1000$ получаем 5,00, это уже с точностью до двух десятичных знаков совпадает с результатом евклидовой геометрии, а следовательно, и с опытом.

В евклидовой геометрии длина единичной окружности — с точностью до четырех десятичных знаков — равна 6,2832. Используя формулу гиперболической геометрии, приходим к следующим результатам:

$$\begin{array}{lll} \text{если} & k = 1, & \bigcirc r = 7{,}3840; \\ \text{если} & k = 10, & \bigcirc r = 6{,}2939; \\ \text{если} & k = 100, & \bigcirc r = 6{,}2832, \end{array}$$

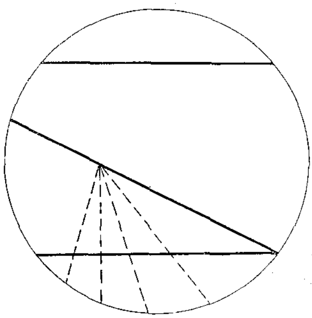
то есть в последнем случае результат с точностью до четырех десятичных знаков совпадает с евклидовым.

Это верно и в общем случае: при $k \to \infty$ метрические соотношения гиперболической геометрии переходят в соответствующие евклидовы соотношения. Таким образом, если выбрать k достаточно большим, то формулы гиперболической геометрии можно точно так же использовать при решении практических (например, технических) задач, как и формулы евклидовой геометрии. Причиной того, что на практике применяются все же евклидовы формулы, является их простота. Иное дело вопрос о том, которая из двух геометрий верна на самом деле. Чему равна в действительности сумма внутренних углов треугольника: 180° или все-таки меньше? Это можно было бы установить лишь с помощью очень точных измерений, однако, пока мы не располагаем достаточно точными измерительными приборами. Согласно общей теории относительности, в пространстве ни евклидова, ни гиперболическая геометрия не являются абсолютно точными, однако в малых масштабах обе они вполне пригодны для описания пространства. С точки зрения упомянутой теории земные масштабы являются еще достаточно «малыми».

КРУГОВАЯ МОДЕЛЬ ГИПЕРБОЛИЧЕСКОЙ ПЛОСКОСТИ КЭЛИ—КЛЕЙНА

Наглядное представление о гиперболической плоскости можно получить при помощи так называемой круговой модели *Кэли—Клейна.*

Точки гиперболической плоскости здесь интерпретируются как точки, лежащие внутри некоторой окружности (так называемой *предельной окружности)*, а прямые — как хорды. Две прямые параллельны, если они пересекаются на предельной окружности, и ультрапараллельны, если соответствующие им хорды не пересекаются. Понятия «принадлежит» и «лежит между» означают то же самое, что и в евклидовой геометрии. Можно доказать, что для модели Кэли – Клейна аксиомы гиперболической плоскости выполняются в том случае, если две фигуры считаются равными, когда существует проективное преобразование круга на себя, переводящее эти фигуры друг в друга. Треугольник, вписанный в предельную окружность, интерпретирует три прямых, каждая из которых параллельна двум другим. Эта необычная фигура действительно существует на гиперболической плоскости — это так называемый *предельный треугольник*, каждая из сторон которого параллельна двум другим.



Параллельные и ультрапараллельные прямые в модели Кэли—Клейна

Использованное здесь понятие модели по существу не отличается от понятия модели в теоретико-множественном смысле (см. «Теория множеств», стр. 574). Отличие лишь в том, что основным элементам гиперболической геометрии на плоскости здесь ставились в соотвествие не теоретико-множественные понятия, а некоторые элементы евклидовой геометрии.

ДИФФЕРЕНЦИАЛЬНАЯ ГЕОМЕТРИЯ

ПРЕДМЕТ ДИФФЕРЕНЦИАЛЬНОЙ ГЕОМЕТРИИ

Изучение кривых и поверхностей более сложной структуры привело к созданию раздела, получившего название *диффереціальной геометрии*. Один из наиболее характерных методов дифференциальной геометрии состоит в том, что фигуры определяются как предельные положения, то есть указываются процессы получения фигур за бесконечное число шагов. Во многих геометрических исследованиях применение такого метода оказывается не только неизбежным, но одновременно необычайно плодотворным, открывая новые пути перед геометрическими исследованиями пространства.

Дифференциальная геометрия в значительной степени опирается на теоремы, модели и приемы математического анализа, отсюда происходит и ее название. Существование исследуемого геометрического объекта здесь связано, как правило, с условиями непрерывности, дифференцируемости или интегрируемости некоторой функции. Для простоты изложения мы заранее предполагаем, что все эти условия выполнены.

ПАРАМЕТРИЧЕСКОЕ ЗАДАНИЕ КРИВЫХ

При рассмотрении параметрических уравнений прямых мы уже видели пример того, как некоторая линия задается в виде функции некоторой вспомогательной переменной. Суть параметрического задания в том, что координаты точек кривой (x, y) на плоскости или (x, y, z) в пространстве задаются как функции некоторого параметра t.

Параметрические уравнения окружности радиуса ϱ с центром в начале координат имеют следующий вид:

$$\boxed{\begin{aligned} x &= \varrho\cos t, \\ y &= \varrho\sin t. \end{aligned}}$$

Когда t принимает вещественные значения из интервала $(0, 2\pi)$, точки с координатами (x, y) описывают всю окружность. Параметр t имеет в этом случае и геометрический смысл: он является направляющим углом вектора с координатами (x, y). (Заметим, что далеко не всегда удается указать простой геометрический смысл параметра.)

Эллипс с большой осью $2a$ и малой осью $2b$ задается в канонической системе координат параметрическими уравнениями

$$\boxed{\begin{aligned} x &= a\cos t, \\ y &= b\sin t. \end{aligned}}$$

Очевидно, что получаемые таким образом x и y действительно удовлетворяют уравнению эллипса $\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$. Если плоская кривая задана уравнением вида $y = f(x)$, то она может быть представлена параметрически с помощью уравнений

$$\boxed{\begin{aligned} x &= t \\ y &= f(t). \end{aligned}}$$

Например, параболу $y = x^2$ можно задать в виде

$$x = t$$
$$y = t^2.$$

Примером параметрического задания кривой в пространстве могут служить уравнения винтовой линии. (Винтовую линию можно получить, если на развертке цилиндра провести прямую, а затем свернуть развертку обратно):

$$x = \varrho \cos t,$$
$$y = \varrho \sin t,$$
$$z = at. \qquad (\varrho,\ a \text{ — постоянные})$$

Кривую можно задать и как линию пересечения двух поверхностей. Например, линия пересечения сферы $x^2 + y^2 + z^2 = \varrho^2$ с цилиндром $x^2 + y^2 = \varrho x$ определяет так называемую *кривую Вивиани*. Ее параметрическое представление имеет вид:

$$x = \varrho \cos^2 t$$
$$y = \varrho \cos t \, \sin t$$
$$z = \varrho \sin t.$$

Если рассмотреть переменный радиус-вектор $\mathbf{r}$, конец которого пробегает всю кривую, то его координаты $\mathbf{r}(x, y\ z)$ являются функциями некоторого параметра t. Таким образом, можно сказать, что сам вектор есть функция некоторого скаляра (числа): $\mathbf{r} = \mathbf{r}(t)$, — так называемая *вектор-функция скалярного аргумента*. Итак, кривую можно задавать вектор-функцией скалярного аргумента. Производную функции $\mathbf{r}(t)$ можно получить, продифференцировав ее координаты. Первая производная функции $\mathbf{r}(t)$ обозначается через $\frac{d\mathbf{r}}{dt}$ или просто $\dot{\mathbf{r}}$ ($\mathbf{r}$ с точкой); вторая и третья производные обозначаются, соотвественно, через $\ddot{\mathbf{r}}$ и $\dddot{\mathbf{r}}$. В заключение заметим, что мы говорим о плоской кривой, если все точки кривой лежат в некоторой плоскости; в противном случае речь идет о пространственной кривой.

ДЛИНА ДУГИ КРИВОЙ; НАТУРАЛЬНЫЙ ПАРАМЕТР

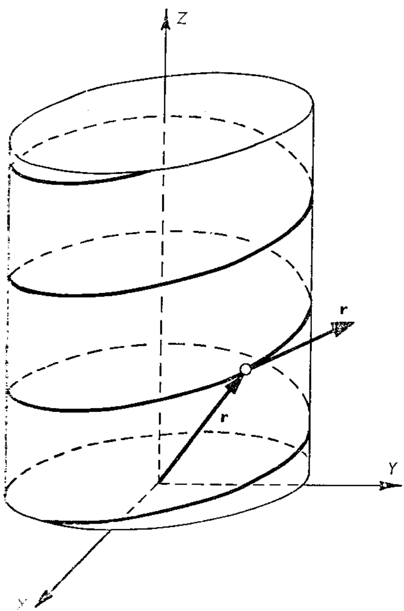
Под длиной дуги кривой понимается наименьшее число (если оно существует), которое больше длины любой ломаной линии, вписанной в эту дугу. Если кривая задана вектор-функцией $\mathbf{r}(t)$, то длина ее дуги, концам которой соотвествуют значения параметра $t = a$ и $t = b$, определяется следующим образом:

$$\boxed{s = \int_a^b |\quad| \, dt.}$$

Если произвольной точке P кривой в качестве параметра ставится в соотвествие длина дуги PO от этой точки до некоторой фиксированной точки O кривой, то такой параметр называется *натуральным*. При задании единичной окружности вектор-функцией $\mathbf{r}$ $(\cos t, \sin t)$ переменная t является натуральным параметром.

КАСАТЕЛЬНАЯ К КРИВОЙ

При исследовании кривой в окрестности некоторой ее точки может возникнуть вопрос о направлении кривой в этой точке. Если некоторое тело под действием некоторой силы движется по криволинейной траектории, то после прекращения действия силы оно продолжит движение уже по прямолинейной траектории.



Винтовая линия и вектор касательной к ней

Под *направлением кривой* в некоторой ее точке P понимается направление прямолинейной траектории, по которой продолжало бы свое движение от точки P некоторое тело, движущееся вдоль этой кривой, если бы сила, действующая на тело, прекратила свое действие как раз в точке P. Геометрически это направление означает направление касательной к кривой в точке P. Ниже будет дано и геометрическое определение касательной.

Если соединить отрезком прямой точку P кривой с произвольной точкой P_1 той же кривой, то мы получим *хорду*. Предельное положение прямой PP_1, когда точка P_1 стремится вдоль кривой к точке P, называется касательной к кривой в точке P. Пусть кривая задана вектор-функцией $\mathbf{r}(t)$, тогда направляющим вектором касательной к этой кривой в точке P_0, соответствующей некоторому значению параметра t_0, есть $\dot{\mathbf{r}}(t_0)$. Единичный вектор направления $\dot{\mathbf{r}}(t_0)$ называется *единичным вектором касательной* или *единичным вектором тангенциального* направления и обозначается через $\mathbf{t}$. Если используется натуральный параметр, то вектор $\dot{\mathbf{r}}(t)$ сам является единичным, в таком случае мы будем обозначать его через $\mathbf{r}'$.

Касательная к винтовой линии образует с осью Z постоянный угол. Действительно, для винтовой линии $\mathbf{r}(t)$ имеем:

$$\dot{\mathbf{r}}(t) \begin{cases} \dot{x} = -\varrho \sin t \\ \dot{y} = \varrho \cos t \\ \dot{z} = a \end{cases}$$

Скалярное произведение вектора $\dot{\mathbf{r}}$ на единичный вектор $\mathbf{k}(0, 0, 1)$ оси Z есть

$$|\dot{\mathbf{r}}|\,|\mathbf{k}| \cos\alpha = a,$$

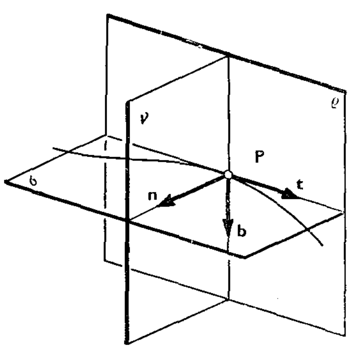
$$\cos\alpha = \frac{a}{|\dot{\mathbf{r}}|\,|\mathbf{k}|} = \frac{a}{\varrho^2 + a^2} = \text{постоянной},$$

то есть α является постоянной.

СОПРОВОЖДАЮЩИЙ ТРЕХГРАННИК

Рассмотрим теперь плоскость, проходящую через касательную к кривой в некоторой точке P и через произвольную точку P_1 кривой. Предельное положение этой плоскости, когда точка P_1 вдоль кривой стремится к точке P, называется соприкасающейся плоскостью кривой в точке P.

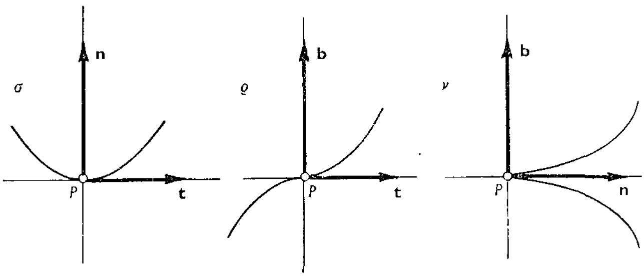
Наглядно это можно выразить так: соприкасающейся является та плоскость, которая лучше всего «приближает» кривую в некоторой окрестности точки P. Естественно, что соприкасающаяся плоскость плоской кривой в любой ее точке совпадает с плоскостью, в которой лежит кривая. Соприкасающаяся плоскость содержит соответствующую касательную к кривой.



Сопровождающий трехгранник пространственной кривой: σ — соприкасающаяся плоскость, ν — нормальная плоскость, ϱ — спрямляющая плоскость

Плоскость, проходящая перпендикулярно касательной к кривой через точку касания, называется нормальной плоскостью кривой в этой точке. Вектор нормали к этой плоскости является направляющим вектором касательной к кривой.

Линию пересечения соприкасающейся и нормальной плоскостей кривой в некоторой точке P называют главной нормалью кривой в этой точке. Построим теперь плоскость, проходящую через точку P перпендикулярно главной нормали. Это так называемая спрямляющая плоскость кривой, линия ее пересечения с нормальной плоскостью называется бинормалью. Введем теперь единичный вектор $\mathbf{n}$ нормали и единичный вектор $\mathbf{b}$ бинормали таким образом, чтобы векторы $\mathbf{t}$, $\mathbf{n}$, $\mathbf{b}$ образовывали правую тройку, то есть чтобы $\mathbf{t} = \mathbf{n} \times \mathbf{b}$. По существу, векторы $\mathbf{t}$, $\mathbf{n}$ и $\mathbf{b}$ задают некоторую прямоугольную систему координат, которая получила название сопровождающего трехгранника.



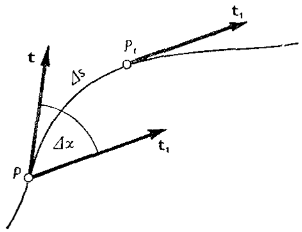
Проекции пространственной кривой на соприкасающуюся, нормальную и спрямляющую плоскости

Исследование поведения кривой в окрестности некоторой ее точки с помощью сопровождающего трехгранника позволяет обнаружить немало интересных свойств. Рассмотрим ортогональные проекции кривой на плоскости сопровождающего трехгранника (предполагается, что в некоторой окрестности точки P кривая является пространственной). Проекция на соприкасающуюся плоскость лежит по одну сторону от касательной и касается ее (эта проекция по форме напоминает параболу). Проекция кривой на спрямляющую плоскость пересекает касательную, а проекция на нормальную плоскость имеет излом в точке P (см. рисунок).

КРИВИЗНА КРИВОЙ, КРУГ КРИВИЗНЫ

Исследуем теперь, насколько отличается кривая в некоторой окрестности точки P от прямой и насколько быстро изменяется ее направление. Поскольку направление кривой в некоторой точке определяется направлением вектора касательной к ней, то изменение направления кривой равносильно изменению направления вектора касательной.

Предположим, что вектор касательной к кривой в точке P в результате перемещения вдоль кривой на Δs перейдет в вектор касательной в точке P_1, изменив при этом направление на угол $\Delta\alpha$. Это изменение характеризуется не только величиной угла поворота, но и тем, насколько «быстро» оно происходит, то есть тем, насколько мала дуга, вдоль которой перемещается вектор. Численно эта характеристика выражается отношением $\frac{\Delta\alpha}{\Delta s}$. Предел этого отношения, когда точка P_1 вдоль кривой стремится к точке P (то есть Δs стремится к нулю), называется *кривизной* кривой в точке P и обозначается через g. В каждой точке прямой кривизна равна 0, и наоборот: если в каждой точке некоторой кривой $g = 0$, то кривая является прямой линией. Кривизна может равняться 0 и в некоторых точках других кривых (например, в точках перегиба). Величина кривизны всегда есть число неотрицательное. Кривизна кривой $\mathbf{r}(t)$ в произвольной ее точке может быть определена по формуле

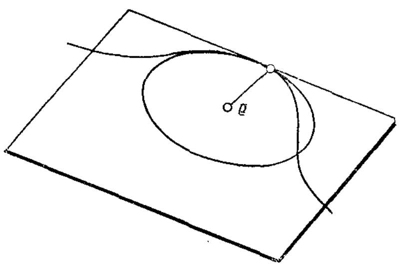


Скорость изменения направления кривой

$$\boxed{g = \frac{|\dot{\mathbf{r}} \times \ddot{\mathbf{r}}|}{|\dot{\mathbf{r}}|^3}.}$$

К понятию кривизны можно прийти и другим путем. Можно сказать, что для изменения направления кривой в некоторой точке P характерно и то, какой дугой окружности можно приблизить эту кривую в достаточно малой окрестности точки P наилучшим образом.

С этой целью выберем окружность, которая касается кривой в точке P (это значит, что в точке P кривая и окружность имеют общую касательную) и проходит через некоторую точку P_1 кривой. Предельное положение такой окружности, когда точка P_1 вдоль кривой стремится к точке P, называется соприкасающейся окружностью или кругом кривизны кривой в точке P.

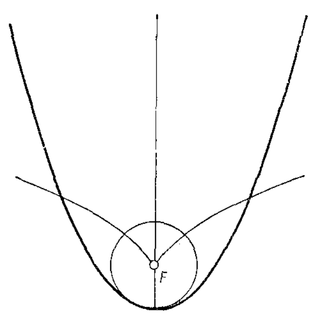


Соприкасающаяся окружность

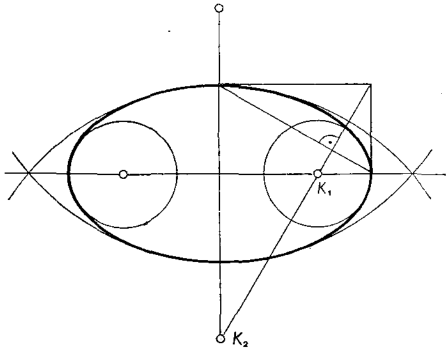
Из самого процесса получения круга кривизны следует, что он лежит в соприкасающейся плоскости кривой и имеет общую касательную с кривой, поэтому его центр лежит на главной нормали. Можно доказать, что величина, обратная радиусу круга кривизны, равна кривизне кривой в данной точке.

Центр круга кривизны называется центром кривизны, соответствующим данной точке.

Центр кривизны играет важную роль при исследовании сил, вызывающих движение тел вдоль данной кривой.



Парабола и ее эволюта



Построение эллипса при помощи соприкасающихся окружностей на концах оси

Для плоской кривой геометрическое место центров кривизны называется эволютой кривой.

Соприкасающаяся окружность плоской кривой, как правило, пересекает кривую в точке касания; иначе обстоит дело в тех точках, через которые проходит ось симметрии некоторого куска кривой. Поскольку соприкасающаяся окружность лучше всех других окружностей приближает кривую в некоторой окрестности рассматриваемой точки, то соприкасающиеся окружности используют и для приближенного построения кривых. На рисунке показан простой способ построения кругов кривизны, соответствующих концам оси эллипса.

Если используется натуральный параметр, то вектор с началом в точке P и концом в центре кривизны есть

$$\boxed{\mathbf{r}'' = g\mathbf{n}\,.}$$

КРУЧЕНИЕ КРИВОЙ

Кривизна пространственной кривой показывает, насколько эта кривая отличается от прямой. Можно поставить также вопрос, насколько пространственная кривая «не является плоской». По аналогии с исследованием кривизны, это можно охарактеризовать скоростью изменения (поворота) соприкасающейся плоскости. Поскольку изменение угла между двумя плоскостями равно изменению угла между векторами нормалей к ним, то рассмотрим угол $\Delta\beta$ между бинормалью кривой в точке P и бинормалью в некоторой точке P_1. Пусть длина дуги PP_1 равняется Δs. Предел отношения $\dfrac{\Delta\beta}{\Delta s}$, когда точка P_1 вдоль кривой стремится к точке P, называется кручением кривой в точке P.

Заметим, что кручение берется со знаком плюс, если при движении точки вдоль кривой в некоторой окрестности точки P в положительном направлении для наблюдателя, смотрящего в отрицательном направлении, единичный вектор бинормали вращается в положительном направлении, и со знаком минус — в противном случае. Таким образом кручение задает скорость изменения направления соприкасающейся плоскости кривой; мы будем обозначать кручение через c.

В каждой точке плоской кривой кручение равно 0, так как в этом случае соприкасающаяся плоскость неизменна. Обратно: если в каждой точке некоторой кривой кручение равно 0, то кривая является плоской. Кручение кривой $\mathbf{r}(t)$ дается формулой

$$\boxed{c = \frac{\dot{\mathbf{r}}\ddot{\mathbf{r}}\dddot{\mathbf{r}}}{|\,\dot{\mathbf{r}}\times\ddot{\mathbf{r}}\,|^2}\,.}$$

ФОРМУЛЫ ФРЕНЕ

Если при задании кривой используется натуральный параметр, то нетрудно вывести основные соотношения между единичными векторами $\mathbf{t}$, $\mathbf{n}$ и $\mathbf{b}$ сопровождающего трехгранника, их производными, а также кривизной и кручением кривой. Это так называемые *формулы Френе:*

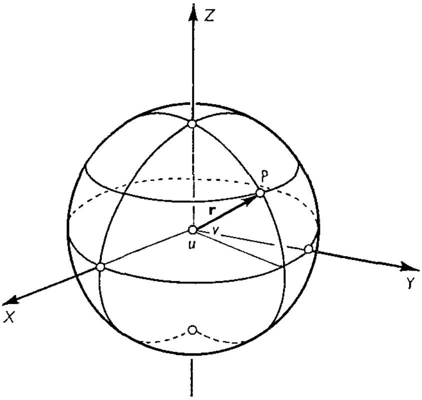
$$\boxed{\begin{aligned} \mathbf{t}' &= g\mathbf{n} \\ \mathbf{n}' &= -g\mathbf{t} + c\mathbf{b} \\ \mathbf{b}' &= -c\mathbf{n}\,. \end{aligned}}$$

ЗАДАНИЕ ПОВЕРХНОСТЕЙ; ГАУССОВЫ КОЭФФИЦИЕНТЫ ПЕРВОГО ПОРЯДКА

Выше мы уже видели немало примеров задания поверхностей. Плоскость, сфера и другие поверхности второго порядка задавались некоторыми уравнениями с тремя неизвестными. Для точек поверхности характерно то, что их координаты удовлетворяют некоторому уравнению с тремя неизвестными, которое в общем случае имеет вид: $f(x, y, z) = 0$.

Если это уравнение можно разрешить относительно z, то есть привести его к виду $z = \varphi(x, y)$, то говорят о задании кривой в форме Эйлера—Монжа

Например, верхняя полусфера, отсекаемая координатной плоскостью XY от сферы радиуса ϱ с центром в начале координат, задается уравнением $x^2 + y^2 + z^2 = \varrho^2$ $(z \geqq 0)$. По методу Эйлера—Монжа эту полусферу можно задать и в виде $z = \sqrt{\varrho^2 - x^2 - y^2}$.



Параметрическое задание сферы

Для геометрических исследований наиболее удобным способом задания поверхностей является двухпараметрический метод Гаусса. Он состоит в том, что задаются радиус-векторы $\mathbf{r}(u, v)$ точек поверхности. В то время, как переменные u и v пробегают все допустимые значения, конец вектора $\mathbf{r}$ описывает всю поверхность. Таким образом, координаты вектора $\mathbf{r}$ являются функциями двух переменных. Если поверхность задана в форме Эйлера—Монжа, то получить ее задание по методу Гаусса можно следующим образом:

$$\boxed{\mathbf{r}(u, v)\begin{cases} x = u \\ y = v \\ z = \varphi(u, v) \end{cases} \quad \text{или} \quad z(u, v) = \big(u, v, \varphi(u, v)\big)\,.}$$

Например: рассмотренная выше полусфера по методу Гаусса задается так:

$$\mathbf{r}(u, v)\begin{cases} x = u \\ y = v \\ z = \sqrt{\varrho^2 - u^2 - v^2}\,. \end{cases}$$

Однако ту же сферу можно задать и при помощи других параметров:

$$\mathbf{r}(u, v)\begin{cases} x = \varrho \sin v \cos u \\ y = \varrho \sin v \sin u \\ z = \varrho \cos v\,. \end{cases}$$

Если пременные u и v являются, в свою очередь, функциями некоторого параметра t, то есть $u = u(t)$, $v = v(t)$, то векторы $\mathbf{r}\big(u(t), v(t)\big) = \mathbf{r}(t)$ описывают на поверхности некоторую кривую. В частном случае, когда $u = t$, $v = c$ (постоянная), вектор $\mathbf{r}$ описывает так называемую u-параметрическую линию. Для сферы эти линии представляют собой или окружности, параллельные координатной плоскости xz (в случае параметров, выбранных в начале) или окружности, параллельные координатной плоскости xy (при втором выборе параметров).

Как мы видели, прямоугольные координаты вектора $\mathbf{r}(u, v)$ являются функциями переменных u и v: $x = x(u, v)$; $y = y(u, v)$; $z = z(u, v)$. Векторы, координаты которых являются частными производными этих функций по u и по v, мы будем обозначать, соответственно, через $\mathbf{r}_u$ и $\mathbf{r}_v$. Точнее: вектор $\mathbf{r}_u$ имеет координаты $\left(\dfrac{\partial x}{\partial u}, \dfrac{\partial y}{\partial u}, \dfrac{\partial z}{\partial u}\right)$, а вектор $\mathbf{r}_v$ — координаты $\left(\dfrac{\partial x}{\partial v}, \dfrac{\partial y}{\partial v}, \dfrac{\partial z}{\partial v}\right)$. Аналогично определяются и векторы $\mathbf{r}_{uu}$, $\mathbf{r}_{uv}$ и $\mathbf{r}_{vv}$ координаты которых являются частными производными второго порядка от соответствующих координат вектора $\mathbf{r}$.

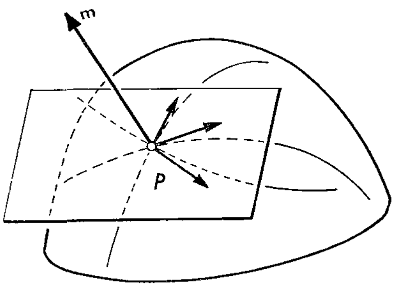
Гауссовыми коэффициентами первого порядка называются следующие величины: $E = \mathbf{r}_u^2$, $F = \mathbf{r}_u \mathbf{r}_v$, $G = \mathbf{r}_v^2$.

Производные функций $u(t)$ и $v(t)$ по переменной t мы будем обозначать, соответственно, через $\dot{u}$ и $\dot{v}$.

КАСАТЕЛЬНАЯ ПЛОСКОСТЬ, НОРМАЛЬ К ПОВЕРХНОСТИ, ГАУССОВЫ КОЭФФИЦИЕНТЫ ВТОРОГО ПОРЯДКА

Исследование поверхностей имеет много общего с исследованием кривых. Особенно важным является то обстоятельство, что изучение поверхностей можно свести к изучению кривых. Поверхность в некоторой точке P можно считать известной, если известны все кривые на этой поверхности, проходящие через точку P и их геометрические свойства.

Расмотрим теперь всевозможные кривые на поверхности, проходящие через точку P, а также касательные к ним в этой точке. Все эти касательные — если только речь идет не об особой точке — лежат в одной плоскости, которая называется касательной плоскостью к данной поверхности в точке P. Вектор $\mathbf{r}_u \times \mathbf{r}_v$ является вектором нормали к этой плоскости; единичный вектор нормали мы будем обозначать через $\mathbf{m}$.



Касательная плоскость и вектор ($\mathbf{m}$) нормали к поверхности

Из сказанного выше следует, что если пересечь поверхность некоторой плоскостью, проходящей через точку P, то касательная к линии пересечения в точке P будет как раз линией пересечения секущей плоскости с касательной плоскостью в точке P.

Величины $L = \mathbf{r}_{uu}\,\mathbf{m}$, $M = \mathbf{r}_{uv}\,\mathbf{m}$ и $N = \mathbf{r}_{vv}\,\mathbf{m}$ называются гауссовыми коэффициентами второго порядка.

ДЛИНА КРИВОЙ НА ПОВЕРХНОСТИ; ПЛОЩАДЬ ПОВЕРХНОСТИ

Длина дуги кривой $\mathbf{r}(u(t), v(t))$ на поверхности между точками, соответствующими значениям параметра $t = a$ и $t = b$, определяется следующим образом:

$$\boxed{\int_a^b \sqrt{E\dot{u}^2 + 2F\dot{u}\dot{v} + G\dot{v}^2}\,dt\,.}$$

Площадь ограниченного куска поверхности можно определить как предел последовательности площадей вписанных в него многоугольников, составленных из треугольников. Однако при этом на треугольники необходимо наложить несколько дополнительных ограничений. Обозначим через T область значений параметров (u, v), соответствующих некоторому куску поверхности $\mathbf{r}(u, v)$, тогда площадь этого куска поверхности равняется

$$\boxed{\iint\limits_T |\mathbf{r}_u \times \mathbf{r}_v| \cdot du\, dv\,.}$$

Через коэффициенты первого порядка это можно выразить в следующем виде:

$$\boxed{\iint\limits_T \sqrt{EG - F^2}\, du\, dv.}$$

КРИВИЗНА КРИВЫХ НА ПОВЕРХНОСТИ

Исследование кривизны кривых на поверхности всегда можно свести к исследованию кривизны плоских сечений поверхности, то есть к изучению кривизны некоторых плоских кривых. Кривизна произвольной кривой на поверхности в некоторой точке P совпадает с кривизной линии пересечения поверхности с соприкасающейся плоскостью этой кривой в точке P. Другими словами, плоскость σ, проходящая через точку P, пересекает поверхность по такой кривой, кривизна которой равна кривизне любой кривой на поверхности, для которой σ является соприкасающейся плоскостью.

Предположим теперь, что некоторая плоскость σ, проходящая через точку P, пересекает поверхность по кривой, кривизна которой равна g, а касательную плоскость к поверхности в точке P — по прямой e. Построим плоскость, проходящую через прямую e перпендикулярно касательной плоскости (так называемую *нормальную плоскость*) и обозначим через α угол между этой плоскостью и плоскостью σ. Пусть кривизна кривой пересечения поверхности с нормальной плоскостью (так называемого *нормального сечения*) равняется g_n. Согласно *теореме Менье*,

$$\boxed{g = \frac{g_n}{\cos \alpha}\,.}$$

Это значит, что кривизна произвольного плоского сечения поверхности может быть определена по кривизне нормального сечения, если только оба они имеют общую касательную. Тем самым исследование кривизны кривых на поверхности сводится к исследованию кривизны нормальных сечений поверхности. Наглядно теорема Менье означает, что если на касательной плоскости к поверхности выбрать некоторую прямую e, проходящую через точку P, а затем вращать вокруг прямой e некоторую плоскость и исследовать получаемые плоские сечения и их кривизну, то наименьшую кривизну в точке P будет иметь нормальное сечение. При этом соприкасающиеся окружности кривых пересечения будут описывать некоторую сферу, а соответствующие им центры кривизны — некоторую окружность.

Благодаря теореме Менье, достаточно рассматривать кривизну нормальных сечений. Заметим, что *кривизна нормального сечения берется со знаком плюс, если направление главной нормали этой кривой совпадает с направлением вектора нормали к поверхности, и со знаком минус — в противном случае.*

Будем теперь вращать некоторую плоскость вокруг нормали к поверхности в точке P. При любом положении эта плоскость высекает из поверхности нормальное сечение.

Обозначим через g_1 минимальную кривизну этих нормальных сечений, а через g_2 — максимальную. g_1 и g_2 называются главными кривизнами поверхности в точке P, соответствующие им кривые — главными сечениями, а направления векторов касательных к этим кривым — главными направлениями.

Если $g_1 \neq g_2$, то главные направления взаимно ортогональны. Рассмотрим теперь произвольное нормальное сечение, кривизну которого обозначим через g, а угол между его плоскостью и главным направлением, соответствующим кривизне g_1 — через φ. Согласно *теореме Эйлера*,

$$\boxed{g = g_1 \cos^2\varphi + g_2 \sin^2\varphi\,,}$$

то есть кривизна всякого нормального сечения может быть определена через главные кривизны.

Векторы $\mathbf{a}$ и $\mathbf{b}$ главных направлений могут быть представлены в виде линейной комбинации векторов $\mathbf{r}_u$ и $\mathbf{r}_v$:

$$\boxed{\mathbf{a} = x_1\mathbf{r}_u + y_1\mathbf{r}_v\,, \qquad \mathbf{b} = x_2\mathbf{r}_u + y_2\mathbf{r}_v\,.}$$

Здесь имеют значение лишь отношения пар коэффициентов (x_1, y_1) и (x_2, y_2); их можно определить из уравнения

$$\begin{vmatrix} y^2 & -xy & x^2 \\ E & F & G \\ L & M & N \end{vmatrix} = 0\,.$$

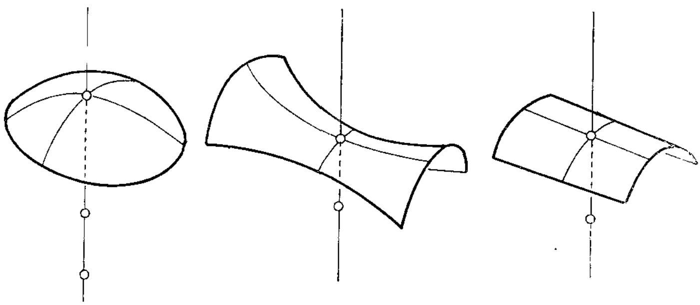
КЛАССИФИКАЦИЯ ТОЧЕК ПОВЕРХНОСТИ; ИНДИКАТОР ДЮПЕНА

Выше мы пришли к заключению, что кривизны в некоторой точке P поверхности известны, если известны главные кривизны g_1 и g_2, то есть g_1 и g_2 определяют кривизны в точке P.

Произведение $K = g_1 g_2$ называется кривизной Гаусса или полной кривизной в точке P; ее можно определить по формуле

$$\boxed{K = \frac{LN - M^2}{EG - F^2}.}$$

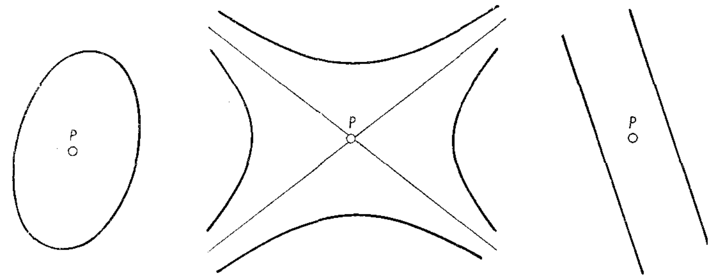
Точки поверхности классифицируют согласно знаку K: если $K > 0$, то точка называется *эллиптической*; если $K = 0$ и одна из величин g_1, g_2 не равна 0, то имеем *параболическую* точку, если же $K < 0$, то точка является *гиперболической*. Эллиптические точки, в которых $g_1 = g_2$, называют *сферическими*, а точки, для которых $g_1 = g_2 = 0$, — *планарными*.



Эллиптическая, гиперболическая, параболическая точки и центры главных кривизн

Эту классификацию точек поверхности хорошо иллюстрирует так называемый индикатор Дюпена. Будем вращать вокруг нормали к поверхности в точке P некоторую плоскость. Если кривая пересечения имеет в точке P кривизну g, то на касательной к ней отмерим в обе стороны от точки P расстояние $\frac{1}{\sqrt{|g|}}$; выполнив это для каждого сечения, мы получим на касательной плоскости некоторую фигуру, которая и называется индикатором Дюпена.

Для эллиптических точек этот индикатор является эллипсом, для гиперболических — состоит из двух гипербол, имеющих общие асимптоты, а для параболических точек — из двух параллельных прямых; для сферических точек получается окружность; для планарных точек индикатора не существует. Для большей наглядности (но менее строго) можно сказать, что если касательную плоскость в точке P немного переместить



Индикатор Дюпена в эллиптической, гиперболической и параболической точках

в направлении поверхности, то отсекаемая ее кривая будет приближением индикатора (для гиперболических точек таким образом можно получить только одну из гипербол).

Все точки произвольного эллипсоида являются эллиптическими или сферическими; все точки сферы — сферические. Точки гиперболического параболоида (седла) и однополостного гиперболоида являются гиперболическими. Для гиперболических точек характерно также, что соответствующие им касательные плоскости пересекают поверхность, причем центры главных кривизн лежат по разные стороны от касательной плоскости. Все точки цилиндра и конуса вращения — параболические. Заметим, что все точки поверхностей, образованных вращением прямой (так называемых *развертывающихся поверхностей*) являются или параболическими, или гиперболическими (если только они не планарны). Всякая точка плоскости является планарной. Важным свойством полной кривизны K является то, что она не меняется, если поверхность изгибать, сохраняя расстояния (то есть таким образом, чтобы длина дуг на поверхности оставалась неизменной). Отсюда следует, например, что на плоскость можно наложить только такую поверхность, в каждой точке которой $K = 0$. По этой причине нельзя, сохраняя расстояния, отобразить сферу на плоскость; поэтому нельзя изготовить карту Земли, сохранив расстояния.

ГЕОМЕТРИЯ НА ПОВЕРХНОСТЯХ; ГЕОДЕЗИЧЕСКИЕ ЛИНИИ

По аналогии со сферической геометрией можно строить геометрии и на других поверхностях. Однако на сфере прямым соответствуют не произвольные линии, а лишь окружности больших кругов; они образуют, например, и стороны сферических треугольников. Окружности больших кругов можно охарактеризовать и как кривые на поверхности, в каждой точке которых нормаль к поверхности совпадает с главной нормалью кривой. Действительно, радиус сферы, являющийся одновременно нормалью к окружности соответствующего большого круга, перпендикулярен к касательной плоскости. Это замечание мы используем при определении так называемых *геодезических линий* — кривых на поверхности, которые соответствуют прямым.

Геодезическими линиями некоторой поверхности называются такие кривые на поверхности, в каждой точке которых нормаль к поверхности совпадает с главной нормалью кривой.

Интересным свойством геодезических линий является то, что кратчайшей дугой (на поверхности), соединяющей две точки поверхности, является геодезическая линия. Обратное утверждение уже не всегда верно: дуга геодезической линии, соединяющая две точки поверхности, не всегда является кратчайшей. Это может происходить всегда, когда геодезическая линия неединственна. Например, две точки сферы, не лежащие на концах одного диаметра, можно соединить двумя геодезическими линиями, одна из которых является кратчайшей из всех дуг, соединяющих эти точки.

Таким образом, при помощи геодезических линий на поверхности можно определять треугольники, различные фигуры и так далее. При этом в зависимости от свойств поверхности мы, естественно, будем получать различные геометрии. С этой точки зрения заслуживают внимания *поверхности постоянной кривизны*, то есть поверхности, на которых значение полной кривизны K постоянно. Если, например, K принимает одно и то же отрицательное значение, то геометрия поверхности является моделью гиперболической геометрии на плоскости. На связь геодезических линий с прямыми указывает и тот факт, что проекция геодезической линии на касательную плоскость к поверхности в некоторой точке P этой линии есть кривая, кривизна которой в точке P равна 0. Наглядно это означает, что в некоторой окрестности точки P эта проекция хорошо приближается отрезком прямой.

Для определения геодезических линий поверхности обычно приходится решать дифференциальные уравнения. Как мы уже отметили, на сфере геодезическими линиями являются окружности больших кругов. На цилиндре вращения геодезическими являются линии, которые при развертке становятся прямыми, то есть образующие цилиндра, винтовые линии и окружности.

ТОПОЛОГИЯ

НЕПРЕРЫВНЫЕ ОТОБРАЖЕНИЯ

При рассмотрении многогранников мы познакомились с теоремой Эйлера, согласно которой число вершин (c), ребер (e) и граней (l) многогранника удовлетворяют соотношению $c + l = e + 2$, т. е.

$$c + l - e = 2.$$

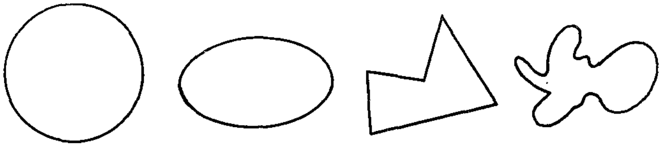
По своему характеру эта теорема существенно отличается от большинства теорем элементарной геометрии — она не содержит метрических соотношений. Еще более очевидным становится ее своеобразие, если заметить, что теорема верна и в более общем случае. В самом деле, представим себе, что рассматриваемый многогранник изготовлен из какого-нибудь эластичного материала, и «раздуем» его, превратив в шар. Вершины и ребра превратяться при этом в узловые точки и соединяющие их линии на поверхности шара, а из граней получатся сферические области; все это будет напоминать глобус. Число узловых точек, ребер и граней на поверхности шара останется тем же, то есть сумма числа вершин и граней будет на 2 больше, чем число ребер. Это число останется неизменным для любой карты на поверхности шара, оно не изменится даже в том случае, если шар продолжить деформировать, но так, чтобы он при этом не «разорвался». Такие деформации мы называем *непрерывными*. Таким образом, число $l + c - e$ не меняется под действием непрерывных деформаций. По существу, непрерывная деформация является отображением, так как точки одной фигуры (тела) она переводит в точки другой. Точнее: отображение называется непрерывным в некоторой точке P, если для сколь угодно малой окрестности K' ее образа P' можно указать такую окрестность K точки P, что образы всех точек из K попадают при этом отображении в K'.

Отображение непрерывно на некотором множестве, если оно непрерывно в каждой точке этого множества. Наглядно непрерывное отображение означает, что при нем не возникает разрывов, так как «достаточно близкие» точки переходят в «достаточно близкие».

Преобразования, с которыми мы познакомились до сих пор — например, преобразования, сохраняющие расстояния, подобные и аффинные преобразования — являются непрерывными, то есть образуют некоторое подмножество множества непрерывных преобразований.

Особенно важное значение имеют взаимно однозначные непрерывные отображения, для которых обратные отображения также являются непрерывными. Такие отображения называются гомеоморфными. Две фигуры (два тела) гомеоморфны, если существует гомеоморфное отображение, переводящее одну из них в другую.

Например, окружности гомеоморфны все фигуры, которые — если изготовить их, например, из резинового жгута — могут быть без разрывов превращены в окружность. Таким образом, окружности гомеоморфны, например, эллипс, контур квадрата и вообще любого простого многоугольника, а также всякая замкнутая ломаная линия без самопересечений в пространстве. Кругу гомеоморфны, например, все простые многоугольники, сегментная поверхность, поверхность куба, лишенная одной из граней. Не гомеоморфны кругу шаровой пояс, сфера, круглое кольцо. Шару гомеоморфны всякий простой многогранник, эллипсоид и так далее.



Фигуры, гомеоморфные окружности

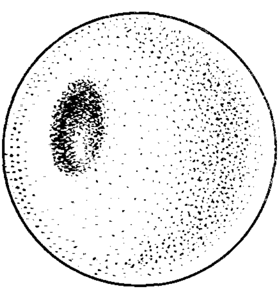
ПРЕДМЕТ ТОПОЛОГИИ

Итак, гомеоморфные отображения переводят грани, вершины и ребра многогранников в такие грани, вершины и линии, для которых сумма $c + l - e$ остается неизменной, то есть является инвариантной. *Топология* занимается свойствами фигур (тел), инвариантными относительно гомеоморфных отображений; сами эти свойства называются *топологическими инвариантами*. При гомеоморфных отображениях топологические инварианты играют такую же роль, какую при преобразованиях, сохраняющих расстояния, играют меры отрезков и углов, а также свойство прямолинейности. С точки зрения топологии ни свойство прямолинейности, ни меры отрезков и углов не являются инвариантными. Ниже мы приведем несколько примеров топологических инвариантов.

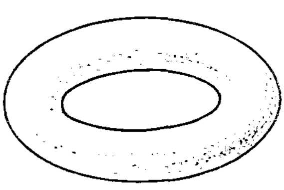
ЭЙЛЕРОВА ХАРАКТЕРИСТИКА ПОВЕРХНОСТЕЙ

Картой на поверхности мы называем такое разбиение поверхности на области, при котором каждая из областей (мы называем их гранями) гомеоморфна кругу. Общие точки линий, ограничивающих области, мы по-прежнему будем называть вершинами, а отрезки линий, соединяющие вершины, — ребрами. Назовем поверхность замкнутой, если она не имеет границы и на нее можно нанести лишь карту, состоящую из конечного числа областей. Полусфера или круг не обладают этим свойством, поэтому они являются открытыми поверхностями.

Пусть карта на некоторой поверхности имеет c вершин, e ребер и l граней. Число $h = c + l - e$ называется эйлеровой характеристикой поверхности.



«Продырявленная» сфера



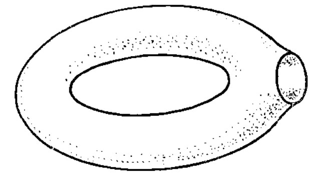
Тор

Эта характеристика не зависит от разбиения поверхности на области, то есть является топологическим инвариантом, характеризующим поверхность (и гомеоморфные ей поверхности). Таким образом для сферы и гомеоморфных ей поверхностей $h = 2$.

Вырежем теперь из карты на сфере одну грань; в результате число граней, а следовательно, и эйлерова характеристика, уменьшится на единицу: $h = 1$. Эта «продырявленная» сфера гомеоморфна кругу, то есть для круга $h = 1$. Вообще, если из некоторой поверхности вырезать область, гомеоморфную кругу, то характеристика Эйлера уменьшится на 1. Итак, эйлерова характеристика круга, а значит, и плоскости, равна 1.

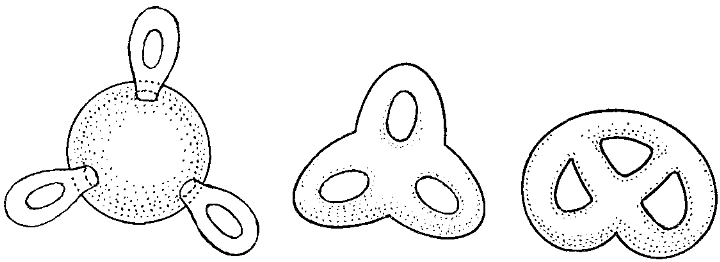
При вращении окружности вокруг не пересекающей ее оси получается поверхность, имеющая форму спасательного круга; она называется *тором*. Эта поверхность гомеоморфна, например, сфере, «просверленной насквозь».

Нарисовав простую карту, нетрудно убедиться, что эйлерова характеристика тора равна 0. Проделаем в торе дырку, гомеоморфную кругу; мы получим поверхность, по форме напоминающую *рукоятку*, ее характеристика равна -1.



Рукоятка

Если в каждой из двух поверхностей имеется дырка, граница которой гомеоморфна окружности, то эти поверхности можно склеить вдоль границ дырок. При этом эйлеровы характеристики поверхностей складываются. Это означает, что если к некоторой поверхности, имеющей дырку, приделать на место дырки рукоятку, то ее характеристика уменьшится на единицу. Таким образом, если в сфере проделать a дырок и, кроме того, приделать к ней указанным выше способом b рукояток (в других местах), то полученная поверхность будет иметь характеристику $2-a-2b$. Изобра-



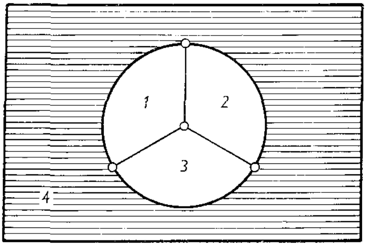
Сфера с тремя рукоятками гомеоморфна «кренделю» с тремя дырками

женные на рисунке «крендели» гомеоморфны сфере с тремя рукоятками, но без дырок, поэтому их эйлерова характеристика равна $2-2\cdot 3=-4$. Сфера, имеющая b рукояток, гомеоморфна поверхности, которая получится, если сферу просверлить насквозь b раз.

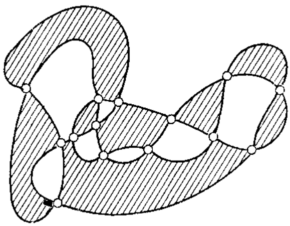
ХРОМАТИЧЕСКОЕ ЧИСЛО ПОВЕРХНОСТИ; РАСКРАШИВАНИЕ КАРТ

Карта называется раскрашенной, если любые две области (грани), имеющие общее ребро, окрашены в разные цвета. Если любая карта на поверхности F может быть раскрашена k красками, но существует карта, которую нельзя раскрасить $k-1$ красками, то k называется хроматическим числом поверхности F.

Нетрудно убедиться, что сфера и плоскость имеют одно и то же хроматическое число. В самом деле, любую карту на плоскости можно получить таким образом, что из некоторой карты на сфере вырезается дырка на



Требуется не менее четырех красок



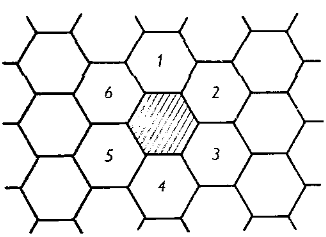
Карта, раскрашенная двумя красками

территории одной страны или внутри области, соответствующей морю, после чего полученная поверхность накладывается на плоскость. Более ста лет назад появилось предположение, что хроматическое число сферы (а значит, и плоскости) равно 4. Это знаменитая *проблема четырех красок*, решить которую пытались многие математики, было испробовано множество методов, однако до нынешнего дня проблема остается нерешенной. Примером того, что для раскрашивания карты нужно, вообще говоря, не меньше 4 красок, может служить круглый остров, изображенный на рисунке. Три радиуса являются границами трех стран. Очевидно, что море, окружающее остров, и каждую из стран необходимо окрасить в разные цвета. Доказано, что хроматическое число плоскости не превышает 5-ти, так как 5-ти цветов достаточно для раскрашивания любой карты на плоскости.

Некоторые задачи, связанные с раскрашиванием карт, решаются проще. Можно показать, например, что для раскрашивания карты, из любой вершины которой выходит четное число ребер, достаточно двух красок. Интересно отметить, что хроматические числа поверхностей, более сложных, чем сфера, уже известны; например, хроматическое число тора равно 7.

ЧИСЛО ПРИКОСНОВЕНИЙ И ПЛОТНОСТЬ

Проблема раскрашивания карт имеет много общего с проблемой плотности фигур (тел). Разобьем некоторую поверхность или кусок поверхности на области.



Число прикосновений мозаики, состоящей из правильных шестиугольников, равно 6

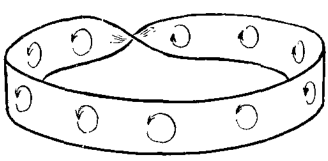
Две области назовем соприкасающимися, если они имеют хотя бы одну точку соприкосновения. Установим для каждой области некоторого разбиения поверхности число соприкасающихся с ней областей; наибольшее из этих чисел назовем числом прикосновений, соответствующим данному разбиению. Если для поверхности (или куска поверхности) существует сколь угодно мелкое разбиение с числом прикосновений s, но не существует произвольно мелкого разбиения с числом прикосновений $s - 1$, то число s называется плотностью поверхности (или куска поверхности).

Нетрудно показать, что для плоскости существует сколь угодно мелкое разбиение с числом прикосновений 6; примером может служить мозаика, составленная из правильных шестиугольников. Следовательно, плотность плоскости не меньше шести. С другой стороны, можно показать, что никакая фигура на плоскости не может иметь плотность, большую шести. Если, например, рассмотреть разбиение единичного квадрата на многоугольники, каждый из которых помещается внутри некоторого круга радиуса 1/30, то среди многоугольников непременно найдется такой, который соприкасается с шестью другими. Таким образом, плоскость и любая фигура на плоскости имеют плотность 6.

НАПРАВЛЕНИЕ НА ПОВЕРХНОСТИ; ОДНОСТОРОННИЕ ПОВЕРХНОСТИ

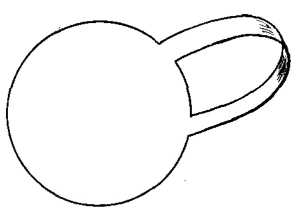
Рассмотрим произвольную окружность (или гомеоморфную ей замкнутую линию) на сфере и установим на ней направление. Как бы мы теперь ни перемещали окружность по сфере, если вернуть ее в начальное положение, то заданное на ней направление не изменится. Этим свойством обладают все поверхности, рассмотренные до сих пор. Говорят, что на этих поверхностях можно установить направление.

Однако этим свойством обладают не все поверхности. Если, например, склеить концы прямоугольной ленты, предварительно перекрутив ее один раз, то мы получим так называемый *лист Мёбиуса*. Нетрудно убедиться в том, что когда перемещаемая по этой ленте направленная окружность вернется в исходное положение, ее направление изменится на противоположное. О таких поверхностях говорят, что *на них нельзя установить направления*. Если бы мы захотели окрасить две стороны листа Мёбиуса в разные цвета, то пришлось бы с удивлением констатировать, что у нее есть только одна сторона — весь лист можно окрасить, не отрывая от нее кисти и не переходя через границу листа.



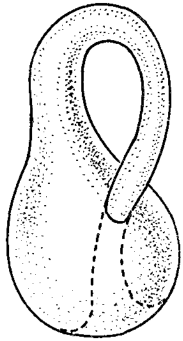
Лист Мёбиуса

Поверхности, на которых нельзя установить направления, называются односторонними, в отличие от поверхностей, на которых можно установить направление, и которые являются двусторонними.



Превращение двусторонней поверхности в одностороннюю путем склеивания

Односторонность и двусторонность являются топологическими инвариантами. Неожиданным свойством листа Мёбиуса является то, что его граница гомеоморфна окружности. Отметим, что эйлерова характеристика листа Мёбиуса равна 0, а хроматическое число равно 6.



Замкнутая односторонняя поверхность Клейна

Открытую двустороннюю поверхность нетрудно превратить в одностороннюю; для этого достаточно приклеить к ее границам концы перекрученной ленты. Существуют, однако, и замкнутые односторонние поверхности. Примером является *поверхность Клейна*. Ее можно получить следующим образом: проделаем в стенке сужающейся трубы отверстие, через которое просунем тонкий конец трубы, после чего соединим края трубы. По всей поверхности Клейна можно пройти, нигде не «проламывая» ее. Эйлерова характеристика этой поверхности равна 0.

Две замкнутые поверхности гомеоморфны, если они имеют одну и ту же эйлерову характеристику и если обе являются или односторонними, или двусторонними.

АБСТРАКТНЫЕ ПРОСТРАНСТВА

Как мы видели, топология занимается лишь немногими, топологически инвариантными свойствами поверхностей, например, сферы. Представляется очевидным, что для аксиоматического построения топологии нет необходимости, например, в аксиомах Гильберта. Действительно, можно указать систему, состоящую из нескольких аксиом, которая фиксирует основное понятие топологии: понятие прикосновения точки к множеству. (Наглядно прикосновение точки к множеству означает, что расстояние между точкой и множеством равно 0, то есть точка лежит бесконечно близко к множеству).

Некоторое множество точек называется топологическим пространством, если для любого его подмножества A известны все точки прикосновения, и выполняются следующие аксиомы:

1. две точки тогда и только тогда соприкасаются, когда совпадают;
2. точка тогда и только тогда является точкой прикосновения объединения двух множеств, когда она является точкой прикосновения хотя бы одного из них;
3. если точка P является точкой прикосновения такого множества, каждая точка которого есть точка прикосновения некоторого множества H, то и P является точкой прикосновения множества H.

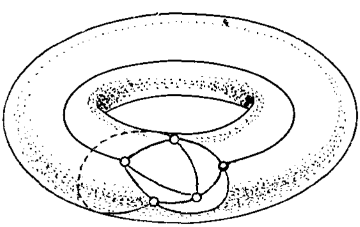
Топологическими свойствами мы называем такие свойства, которые основываются на понятии прикосновения точки к множеству. Таким образом, топология является разделом геометрии, изучающим топологические свойства геометрических объектов. Основываясь на понятии прикосновения, непрерывные отображения можно определить как отображения, сохраняющие свойство прикосновения.

Топологию принято делить на два больших раздела. В *теоретико-множественной топологии* применяют главным образом методы теории множеств и теории функций. В *комбинаторной* (или *алгебраической*) *топологии* используется аппарат комбинаторики и алгебры (например, теории групп).

ТЕОРИЯ ГРАФОВ

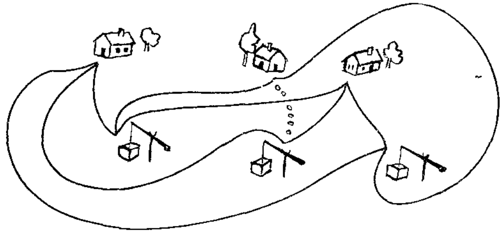
ВОЗМОЖНОСТЬ НАНЕСЕНИЯ КАРТ НА ПЛОСКОСТЬ

Интересным свойством тора является и то, что на нем можно нарисовать карту, имеющую пять вершин, каждые две из которых соединяют одно ребро. Это свойство оказывается неожиданным, так как после нескольких попыток нетрудно убедиться, что на плоскости такую карту нарисовать нельзя: как бы мы ни начали построение, всякий раз получается, что последнее ребро должно пересечь несколько предыдущих. Не удается решить на плоскости и задачу-шутку «три дома — три колодца», которая заключается в следующем: на некотором (плоском) участке построены три дома и три колодца; требуется провести пути от каждого дома к каждому колодцу таким образом, чтобы они не пересекались. Докажем, что это невозможно. В самом деле, если бы такие пути существовали, то это была бы карта с шестью вершинами ($c = 6$), число ребер которой было бы $e = 3 \cdot 3 = 9$. Каждая грань была бы ограничена четным числом ребер, так как один конец всякого ребра находится у дома, а другой — у колодца, и, значит, любой замкнутый путь должен состоять из четного числа ребер. Таким образом, всякий замкнутый путь состоит по крайней мере из четырех ребер. Эйлерова характеристика поверхности равна 1, поэтому $6 + l - 9 = 1$, то есть карта должна состоять из 4 граней; обозначим число ограничивающих их ребер, соответственно, через n_1, n_2, n_3 и n_4, а число ребер, составляющих границу всей карты — через v; (v также не может быть меньше 4-х, так как граница сама является замкнутым путем). В результате получается, что сумма $n_1 + n_2 + n_3 + n_4 + v$ должна быть не меньше 20, в то же время эта сумма должна ровно вдвое превышать число ребер, то есть должна равняться 18. Отсюда следует, что искомой карты не существует.



Карта на торе, которую нельзя перенести на плоскость

После этого, естественно, возникает вопрос, каким условиям должна удовлетворять некоторая карта, чтобы ее можно было нанести на плоскость. Ответом является то, что по существу, нанесению произвольной карты на плоскость препятствуют карты двух указанных типов. Если карта не содержит участка, строение которого совпадает или с рассмотренным выше пятиугольником, или с фигурой из задачи «три дома — три колодца», то эту карту можно нанести на плоскость.



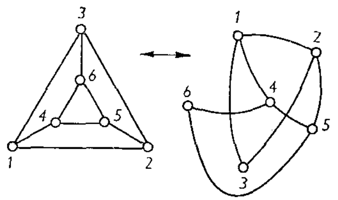
Задача «три дома — три колодца»

ПОНЯТИЕ ГРАФА; ОСНОВНЫЕ ТЕРМИНЫ

Рассмотренная выше карта является фигурой, состоящей из вершин и ребер. Такие фигуры называются графами.

Ребра графов обычно изображают непрерывными линиями. Обычно допускается также, чтобы ребра графа пересекались не только в вершинах, однако, такие точки пересечения не считаются вершинами, поэтому на рисунках мы выделяем вершины графа особо.

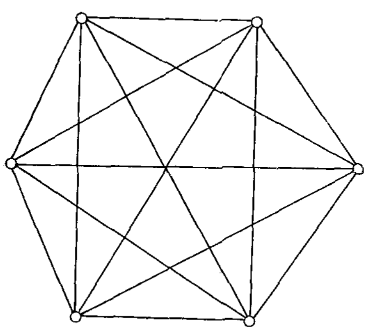
Два графа называются изоморфными, если между их вершинами можно установить такое взаимно однозначное соответствие, что две вершины одного графа тогда и только тогда соединены ребром, когда соединены и соответствующие им вершины другого графа.



Пример изоморфных графов. Соответствующими являются вершины, имеющие одинаковые номера

Если в графе, имеющем n вершин, любые две вершины соединены ребром, то граф называется полным графом с n вершинами.

В полном графе с n вершинами каждая вершина является точкой пересечения $n-1$ ребер; общее число его ребер равно $\frac{n(n-1)}{2}$.

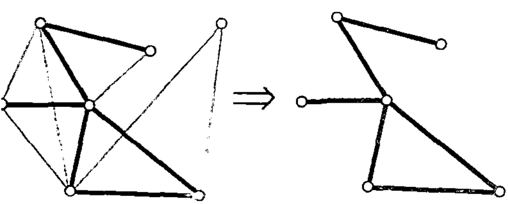


Полный граф с шестью вершинами

Подграфом G' графа G мы называем всякий граф, вершины и ребра которого являются, соответственно, вершинами и ребрами графа G.

Говорят, что граф можно изобразить на плоскости, если существует изоморфный ему граф, ребра которого не имеют иных точек пересечения, кроме вершин.

Граф из задачи «три дома — три колодца» нельзя изобразить на плоскости, однако он, разумеется, существует.



Граф и его подграф

Используя введенные термины, можно сказать, что некоторый граф нельзя изобразить на плоскости, если у него существует подграф, изоморфный или графу из задачи «три дома — три колодца», или полному графу с пятью вершинами. Заметим, что если граф можно изобразить на плоскости, то существует изоморфный ему граф, ребра которого являются отрезками прямых.

Если не оговорено противное, то мы будем считать, что произвольный граф имеет конечное число вершин, и любые две вершины могут быть соединены не более, чем одним ребром.

СТЕПЕНЬ

Степенью произвольной вершины графа называется число ребер, сходящихся в этой вершине.

Число нечетных вершин (вершин с нечетной степенью) любого графа четно. Для доказательства этого утверждения обозначим степени четных вершин графа через $a_1, a_2, \ldots, a_k$, а степени нечетных вершин — через $b_1, b_2, \ldots, b_r$. Сумма

$$a_1 + a_2 + \ldots + a_k + b_1 + b_2 + \ldots + b_r$$

ровно вдвое превышает число ребер графа (каждое ребро мы подсчитали дважды), то есть является четным числом. Сумма $a_1 + a_2 + \ldots + a_k$ также является четной (как сумма четных чисел), а следовательно, и сумма $b_1 + b_2 + \ldots + b_r$ должна быть четной. Это возможно лишь в том случае, если r четно, то есть число нечетных вершин четно. Доказанное только что утверждение можно сформулировать многими способами. Например: в любой компании число людей, у которых в этой компании нечетное число знакомых, всегда является четным. Действительно, всех членов компании можно изобразить точками и соединять две точки лишь в том случае, если соответствующие им люди знакомы друг с другом; тогда число знакомых каждого человека будет равно степени соответствующей вершины графа.

Аналогичным образом можно показать, что этой теореме теории графов равносильны и следующие утверждения:

Число живущих и живших людей, которые до сих пор пожимали руку другим людям нечетное число раз, является четным.

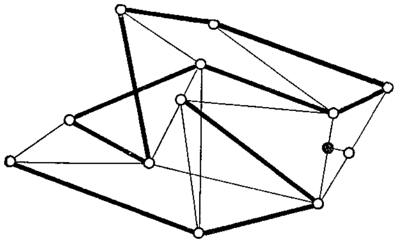
Число вершин многогранника, в которых сходится нечетное число ребер, четно.

Граф, степени всех вершин которого равны k, называется однородным графом степени k.

Примером однородного графа степени 3 является граф, состоящий из ребер куба; однородным графом степени 4 является граф, состоящий из ребер октаэдра.

ЦЕПИ, ЦИКЛЫ; СВЯЗНЫЙ ГРАФ; ДОПОЛНИТЕЛЬНЫЙ ГРАФ

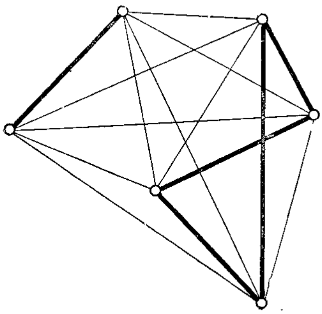
Отметим на некотором графе последовательность соединенных друг с другом ребер таким образом, чтобы два соседних ребра последовательности имели общую вершину, и никакое ребро не присутствовало в последовательности дважды. Если в этой последовательности никакие три ребра не имеют общей точки (то есть последовательность не является самопересекающейся), то она называется цепью; если начальная и конечная точка цепи совпадают (замкнутая цепь), то ее называют циклом. Длиной цепи или цикла называется число составляющих их ребер.



Цепь и цикл на графе

Граф называется связным, если любые две его вершины соединяет некоторая цепь.

Дополнением графа G называется такой граф $\overline{G}$, вершины которого совпадают с вершинами графа G, а ребрами являются те ребра, которые следует добавить к графу G, чтобы получился полный граф.



Дополнительный граф

Если некоторый граф не является связным, то связно его дополнение, и наоборот (может случиться, что оба графа связны). Если, например, в какой-нибудь области автобусное сообщение между населенными пунктами обслуживают две транспортные фирмы, причем любые два населенных пункта связывает маршрут автобуса одной из фирм (и только одной), то по крайней мере одна из фирм такова, что, пользуясь ее автобусами, можно из любого пункта попасть в любой другой.

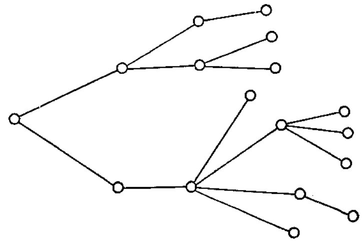
(Вершинами графа здесь являются населенные пункты, а ребрами — маршруты автобусов одной из фирм; маршруты автобусов другой фирмы являются ребрам дополнения.)

ДЕРЕВЬЯ

Деревом называется связный граф, не имеющий циклов.

Своим названием этот граф обязан тому, что его изображение напоминает разветвленное дерево.

Наглядное представление о числе ребер дерева с вершинами можно получить следующим образом: будем поочередно «обламывать» ребра дерева, начав с концов ветвей; вместе с каждым ребром мы будем убирать и одну вершину; в конце останется одна вершина, то есть дерево имеет $n-1$ ребер. Если некоторое число населенных пунктов хотят связать телефонными линиями так, чтобы из любого пункта можно было связаться по телефону с любым другим пунктом, причем затраты должны быть минимальными, то очевидно, что граф телефонной сети должен быть деревом. Действительно, если бы этот граф содержал некоторый цикл, то это означало бы, что построена по крайней мере одна лишняя телефонная линия, так как удаление одного из ребер цикла еще гарантирует связь между всеми его вершинами. Если известны затраты на постройку телефонных линий между отдельными пунктами, то можно указать метод построения графа, обеспечивающего минимальную сумму затрат.



Дерево

ЭЙЛЕРОВА И ГАМИЛЬТОНОВА ЛИНИИ ГРАФА

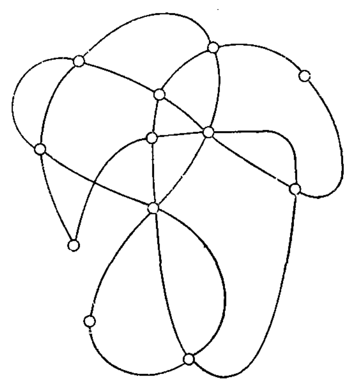
Одной из наиболее старых задач теории графов является вопрос о возможности изобразить граф единственной замкнутой линией.

Замкнутая последовательность соединенных друг с другом ребер, которая содержит каждое ребро графа ровно один раз, называется эйлеровой линией графа.

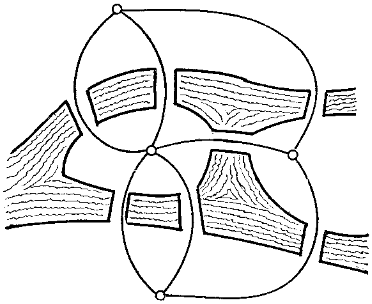
Условие существования такой линии впервые было указано *Эйлером*. Заметим, что речь идет, естественно, о связном графе, причем здесь мы в виде исключения допустим, чтобы две вершины связывало несколько ребер.

Если граф является *эйлеровым* (то есть содержит эйлерову линию), то, двигаясь вдоль этой линии и входя в некоторую вершину по некоторому ребру, мы должны выйти из нее по некоторому другому ребру, то есть в каждой вершине должно быть четное число ребер, а значит, все вершины являются четными. С другой стороны, можно показать, что это условие является достаточным для существования эйлеровой линии. Таким образом, связный граф тогда и только тогда может быть изображен в виде одной замкнутой линии, когда все его вершины являются четными.

Если граф содержит только две нечетные вершины, то и в этом случае его можно изобразить одной линией, но она уже не будет замкнутой; нечетные вершины будут концами этой линии. Вообще:



Эйлеров граф



Кёнигсбергские мосты

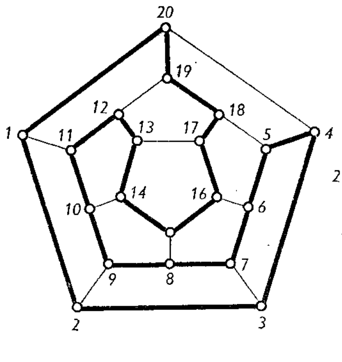
если на связном графе имеется $2k$ нечетных вершин, то граф может быть изображен k открытыми линиями, которые в совокупности содержат все ребра графа в точности по одному разу.

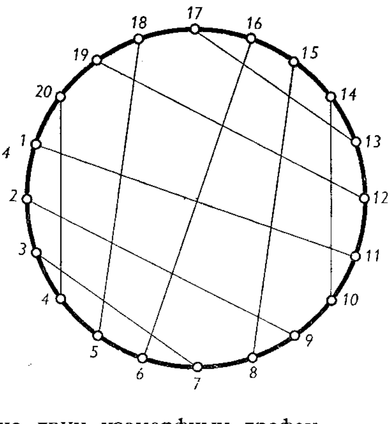
«Родоначальницей» всех задач этого типа явилась так называемая *задача о кёнигсбергских мостах*. Город Кёнигсберг (ныне Калининград) расположен на берегах реки Прегель и на двух островах, которые были связаны несколькими мостами. Эйлеру был задан вопрос, можно ли пройти по всем мостам, перейдя через каждый в точности один раз. Пусть четыре вершины некоторого графа изображают берега реки и два острова, а ребра — связывающие их мосты. Поскольку все вершины графа нечетные, то, согласно предыдущей теореме, искомого пути не существует, то есть граф нельзя изобразить одной линией.

Проблема существования эйлеровой линии несколько напоминает проблему существования линии Гамильтона. Гамильтоновой линией называется цикл, проходящий по всем вершинам графа. Граф называется гамильтоновым, если он содержит гамильтонову линию.

Наглядно такой граф можно изобразить, расположив его вершины по некоторой окружности. Для существования гамильтоновой линии уже нельзя указать такого простого критерия, как для эйлеровой линии; справедливо, однако, утверждение, что если степень всякой вершины графа, имеющего n вершин ($n \geqq 3$), не меньше $\frac{n}{2}$, то у него существует линия Гамильтона.

Используя упомянутую выше модель компании, можно сказать, что если в компании из n членов ($n \geqq 3$) каждый человек имеет по крайней





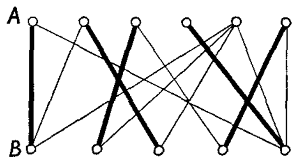
Гамильтонова линия на двух изоморфных графах мере $n/2$ знакомых, то всю компанию можно усадить вокруг стола таким образом, что каждый из сидящих будет знаком с двумя своими соседями по столу. Вершины графа соответствуют членам компании, и две вершины соединены, если соответствующие им люди знакомы друг с другом.

Заметим, что данное условие является достаточным, но не необходимым для того, чтобы граф был гамильтоновым. Например, для гамильтонова графа, изображенного на рисунке, это условие не выполняется.

ЧЕТНЫЕ ГРАФЫ

Из приведенных выше примеров видно, что графы могут быть использованы для наглядной иллюстрации связей между некоторыми объектами. Объектам соответствуют вершины, а связям — ребра графа. Может случиться, однако, что имеются объекты двух видов, и их необходимо различать. Например, биржа труда предлагает a возможностей работы; в то же время имеется b работников, которые готовы выбрать одну или, может быть, несколько возможностей. Чтобы проиллюстрировать эту зависимость с помощью графа, необходимо разделить его вершины на две группы. К группе A относятся вершины, соответствующие a возможностям работы, а к группе B — вершины, соответствующие работникам. Если некоторый работник готов взяться за какую-нибудь работу, то соответствующие вершины графа соединяются. В результате мы получим граф, вершины которого можно разделить на две непересекающиеся группы A и B; ребра могут соединять лишь вершины, принадлежащие разным группам. Такие графы называются четными. Характерным свойством четных графов является то, что всякий цикл на них состоит из четного числа ребер, поэтому такие графы называют также *графами с четным обходом*.

В связи с рассмотренной выше моделью распределения работ возникает вопрос, при каких условиях каждый работник может получить нужную ему работу. Одно из таких благоприятных распределений работ между работниками иллюстрирует на рисунке так называемая *система независимых ребер*, покрывающая A, то есть такое множество ребер, никакие два из которых не имеют общего конца, и в то же время из каждой вершины группы A выходит некоторое ребро этого множества. Можно доказать, что если при любом выборе k вершин из A они в совокупности соединены не менее, чем с k вершинами из B, то существует система независимых ребер, покрывающая A, и задача распределения работ имеет решение. (Заметим, что условие должно выполняться при любом k; практически это означает, что все k работников вместе должны быть готовы к выполнению по крайней мере k работ, что, очевидно, является необходимым условием разрешимости задачи.)



Четный граф (жирными линиями выделена система независимых ребер)

Если четный граф является однородным, то множества A и B содержат одинаковое число вершин. В этом случае всегда существует система независимых ребер, покрывающая *все* вершины.

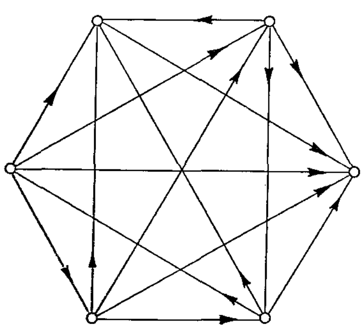
Если, например, некоторая газета объявляет конкурс на решение n кроссвордов, на что откликается n читателей, причем каждый из них посылает решения k кроссвордов ($k \leq n$), и на каждый кроссворд приходит k решений, то редакция газеты может огласить результаты конкурса таким образом, что сообщает о решении разных кроссвордов разными читателями. Вершинами графа здесь являются кроссворды и читатели, а ребра соответствуют решениям.

ОРИЕНТИРОВАННЫЕ ГРАФЫ

В рассмотренных выше примерах связи, изображавшиеся ребрами графов, носили одинаковый характер. Если, например, для проведения шахматного турнира мы хотим указать, какие шахматисты с кем играют, то решение дается очень простым графом: вершины изображают шахматистов, а соединяющие их ребра означают, что соответствующие шахматисты играли друг с другом. Однако в таком графе не отражается результат игры. Предположим, что ничьих не было; в этом случае победителя можно обозначить, *установив направление* на соответствующем ребре графа (на рисунках это обозначается стрелкой). Граф, на котором указаны направления ребер, называется *ориентированным*.

В связи с ориентированными графами также возникает немало интересных проблем. Например, ориентированный граф тогда и только тогда содержит (замкнутую) эйлерову линию, когда для каждой его вершины число ребер, направленных к ней, совпадает с числом ребер, направленных от нее. Под направленной линией мы понимаем такую линию на графе, на которой соседние ребра имеют одинаковое направление.

Интересным свойством обладают полные ориентированные графы: у них всегда имеется вершина, из которой в любую другую вершину можно попасть по направленной цепи, состоящей максимум из двух ребер. Турнир, в котором каждый сыграл с каждым один раз, причем ничьих не было, можно проиллюстрировать полным ориентированным графом. Высказанное выше утверждение в этом случае означает, что среди участников турнира найдется такой, который назовет имена всех участников турнира, если укажет всех тех шахматистов, у которых он выиграл (с соответствующими вершинами его соединяют направленные ребра) и тех, у которых выиграли проигравшие ему (с соответствующими вершинами его соединяют направленные цепи длины 2).



Полный ориентированный граф

ОБОБЩЕНИЕ ПОНЯТИЯ ГРАФА

Понятие графа возникло из геометрии, но, как видно из приведенных примеров, в большинстве задач и теорем теории графов геометрия служит лишь для иллюстрации, в чем нередко и заключается основная роль графов. Однако с течением времени понятие графа стало самостоятельным. Не используя основных понятий геометрии, граф можно определить следующим образом: произвольное множество считается графом, если между его элементами определена некоторая связь, причем о любых двух элементах известно, существует ли между ними эта связь.

Можно дать и такую формулировку: некоторое множество называется графом, если на этом множестве определена некоторая функция $f(x, y)$ двух переменных так, что для любых двух элементов a и b ($a \neq b$) этого множества $f(a, b) = f(b, a) = 0$ или $f(a, b) = f(b, a) = 1$. Значение 0 означает, что a и b не связаны, а 1 — что a и b соединены ребром.

В этих определениях мы не требовали, чтобы множество было конечным, то есть они могут определять и *бесконечный граф*. Простой пример бесконечного графа можно получить, если считать вершинами целочисленные точки двух числовых прямых и соединить ребрами точки, соответствующие одному и тому же числу.

КРАТКИЙ ОЧЕРК ИСТОРИИ ГЕОМЕТРИИ

Геометрия — одна из наиболее древних и ранее других систематизированная ветвь математики. Первые ее следы можно обнаружить у египтян, занимавшихся земледелием. Задачи, возникавшие в связи с измерением земельных участков, привели их к осознанию основных положений геометрии. Измерение земельных участков было необходимо, так как ежегодные разливы реки Нил смывали границы участков, которые таким образом нужно было периодически восстанавливать. От египтян основы геометрии узнали греческие философы; среди них следует отметить Талия (639—548 г. до н. э.), который долгое время изучал в Александрии достижения египетской науки, а затем возвратился в Грецию. Считают, что именно он ввел понятие подобия треугольников и измерение углов дугами. Пифагор (580—500 г. до н. э.) очевидно познакомился с частным случаем теоремы, названной его именем, во время своих путешествий, так как египтяне, халдеи и индусы безусловно знали, что треугольник со сторонами 3, 4, 5 является прямоугольным. В общем виде теорема была высказана и доказана вероятно самим Пифагором. Его школа подняла геометрию до уровня науки.

IV—III века до н. э. явились периодом расцвета греческой геометрии. Центром греческой школы была Александрия, а выдающимися ее представителями — Евклид, Архимед и Аполлоний. Евклид (315—255 г. до н. э.) в своем 13-томном труде под названием «*Начала*» систематизировал геометрию. Наиболее значительным результатом Архимеда (287—212 г. до н. э.) в геометрии была разработка процесса квадратуры и методов ее применения к вычислению площадей и объемов, ограниченных кривыми и поверхностями. Аполлоний (жил, вероятно, между 260 и 200 годами до н. э.) заложил основы теории плоских сечений конуса. Он исследовал их методами алгебры, и поэтому может считаться одним из предвестников аналитической геометрии.

Во II веке нашей эры образовалась новая александрийская школа. Менелай (жил около 100 года) основал теорию сферических треугольников. Птолемей (около 150 года) в своем рукописном труде по астрономии приводит много геометрических теорем и делает обзор геометрии своего времени. Он составил таблицу длин хорд окружности до 90° с шагом 30. Это первые таблицы синусов. Современником Менелая был Герон, который решал геометрические задачи путем вычислений. Еще более видным представителем этой школы является Папий (IV век н. э.). Он собрал воедино результаты греческой геометрии, многие из которых дошли до нас именно от него: Папий занимался и определением объемов некоторых тел вращения, а также задачами на минимум и максимум.

В V—VI веках в развитии геометрии наблюдается застой, появляются лишь работы исторического характера. В Азии, основываясь на результатах греков, занимаются главным образом измерениями и тригонометрией. В Индии элементами геометрии становятся правила постройки алтарей. Занимаются определением числа π. Задачи, связанные с измерениями, встречаются и у так называемых 10 китайских классиков. Они вычисляли π с помощью периметров вписанных и описанных многоугольников. Из арабских математиков Албатеги (877—929) занимался тригонометрией, Абул Вефа ввел понятие тангенса и котангенса; Алхазен (965—1039), следуя Архимеду, вычислил объем тела, образованного вращением параболы. Персидский поэт Омар Хаям (около 1123) производил вычисления, связанные с плоскими сечениями конуса. Севильский математик Ябир ибн Афлах (жил в первой половине XII века) развивает теорию сферических треугольников. Нашир-ад-Дин (1201—1274) из Багдада пробует доказать аксиому параллельности Евклида. В это время тригонометрия становится независимой от астрономии. В 1220 году Леонардо Пизанский в своем труде под названием Practica geometriae излагает геометрию своего времени. Нюрнбергский астроном Региомонтанус (Йоганнес Мюллер, 1436—1476) пишет первый труд, посвященный только тригонометрии.

В XV веке великие художники Возрождения своими исследованиями законов перспективы обогащают знания о пространстве. Леонардо да Винчи занимался, кроме того, построениями эллипса, а А. Дюрер способствовал развитию геометрии изучением геометрических преобразований и своими указаниями к использованию циркуля.

В 1596 году Лудольф ван Цейлен определяет число π с точностью до 20-ти, а затем до 35-ти десятичных знаков. Ф. Виет (1540—1603) совершенствует тригонометрию. Необходимость в изготовлении карт, появившаяся с развитием торговли и мореплавания, привела к развитию теории проекций; была открыта проекция Меркатора. Португалец П. Нуньез (1492—1577) открывает локсодромы шара.

Появляются новые методы вычисления площадей, объемов и определения центра тяжести. С. Стевин (1548—1620), Й. Кеплер (1571—1630) и Б. Кавальери (1598—1647) закладывают основы интегрального исчисления.

П. Ферма (1601—1665) указывает методы нахождения максимумов и минимумов кривых. Изучаются новые кривые, например, Б. Паскаль (1623—1662) изучал циклоиду, введенную Галилеем.

Р. Декарт (1596—1650) применяет к греческой геометрии алгебраические методы и тем самым создает аналитическую геометрию. Независимо от него к тем же результатам приходит и Ферма. Его друг Ж. Дезарг (1591—1661) ведет исследования в прямо противоположном направлении и в своей книге, появившейся в 1639 году, излагает основы проективной геометрии. Я. де Витт и Дж. Уоллис применяют методы аналитической геометрии к плоским сечениям конуса. Уоллис находит представление числа π бесконечной последовательностью, В. Браункер — бесконечной цепной дробью, а Дж. Грегори и Г. В. Лейбниц — бесконечным рядом. И. Барроу в работе, вышедшей в 1668 году, показывает, что определение направления касательной и нахождение площади под кривой являются взаимно обратными задачами. Так на основе аналитической геометрии начинают вырисовываться контуры дифференциального и интегрального исчисления. X. Гейгенс занимается изучением эволют и эвольвент плоских кривых. Паскаль, основываясь на теореме Дезарга, устанавливает свою теорему о плоских сечениях конуса.

Исследования Ньютона в геометрии связаны с кривыми третьего порядка. Г. Лопиталь опубликовал в 1707 году свое исследование плоских сечений конуса. Братья Бернулли систематически использовали дифференциальное и интегральное исчисление для определения площадей фигур, ограниченных кривыми.

В 1735 и 1780 годах Л. Эйлер публикует свои работы по дифференциальной геометрии. Он открыл и прямую в треугольнике, носящую его имя.

Видным геометром своего времени был Ж. Роберваль (1602—1675). Он ввел, например, и понятие параллелограмма векторов. К. Маклорин, де Гуа и Г. Крамер занимались изучением алгебраических кривых. Англичанин Т. Симпсон (1710—1761) в своей книге, посвященной тригонометрии, дает простую формулу для вычисления значений синуса и косинуса. В 1770 году появляется книга Ж. Ламбера, в которой он исследует гиперболические синус и косинус; он же ввел и названия этих функций. Примечательны также его работы по теории перспективных проекций. Г. Монж (1746—1818) посвятил геометрии два труда; в них он указывает новые, чрезвычайно важные пути развития технических применений геометрии. Важное значение имеют и его работы по дифференциальной геометрии. Его ученик Ж. Понселе (1788—1867) продолжает развитие проективной геометрии. Й. Штейнер (1796—1863) создает синтетическую геометрию. К. Штаудт (1798—1867) строит новую геометрию, не использующую метрических соотношений, и тем самым становится одним из основоположников топологии. Профессор боннского университета Ю. Плюккер (1801—1868) применяет алгебраические методы к исследованию алгебраических кривых и поверхностей. Учеником Монжа был Ш. Дюпен (1784—1873), который продолжил развитие дифференциальной геометрии и ввел понятие индикатора, носящего его имя, а также другие новые понятия.

К. Ф. Гаусс (1777—1855) сделал очень много и в геометрии. В своей книге, появившейся в 1827 году, помимо изложения результатов, касающихся элементарной геометрии, он ввел новые важные понятия в теории поверхностей, сразу ставшие классическими. Гаусс занимался и аксиомой параллельности Евклида.

Открытие абсолютной геометрии является заслугой венгерского математика Яноша Бойаи (1802—1860) и русского математика Н. И. Лобачевского (1793—1856), которые, будучи современниками, независимо друг от друга разработали неевклидову геометрическую систему. Г. Грассман заложил основы векторной алгебры и тензорного исчисления. В 1854 году результаты Гаусса в дифференциальной геометрии привели Б. Римана (1826—1866) к теории многослойных поверхностей и к признанию значения топологии.

Систематизацию новых геометрических понятий произвел Ф. Клейн (1849—1925) в своем труде под названием «Эрлангенской программы» (1872). Работы Л. Чивита (1873—1941) и Э. Картана (1910) привели к созданию новых геометрических систем.

Развитие геометрии в XIX и XX веках привела к необходимости исследовать *основы* геометрии. Эту работу выполнил немецкий математик Д. Гильберт (1862—1943) в своих фунндаментальных трудах, первый из которых вышел в свет в 1899 году.

МАТЕМАТИЧЕСКИЙ АНАЛИЗ

ВВЕДЕНИЕ

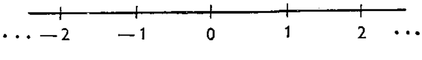
Математический анализ занимается вопросами, связанными с понятием *предела*. Немало таких понятий встречается в курсе средней школы (например, длина окружности и площадь круга, площадь сферической поверхности и объем шара, иррациональное число и др.). Более того, поскольку в результате общественного прогресса все более широко используется научная терминология, такие понятия встречаются и в повседневном словоупотреблении (скорость, ускорение, непрерывная кривая, плотность и др.). В дальнейшем мы увидим, что эти понятия не могут быть определены удовлетворительно без использования понятия предела.

Два типа пределов играют основополагающую роль как в самой математике, так и в приложениях. Это *производная* и *интеграл*. Оба эти предела появились уже в древнегреческой математике: в задаче Аполлония о касательных и в работах Архимеда, посвященных вычислению объемов. Однако определение этих понятий в общем виде и установление связи между этими двумя важнейшими типами пределов связано с именами Ньютона и Лейбница и является крупнейшим достижением науки XVII века. С тех пор и говорят о «дифференциальном и интегральном исчислении», ставшем основой математического анализа и незаменимым вспомогательным аппаратом для математической трактовки естественных наук.

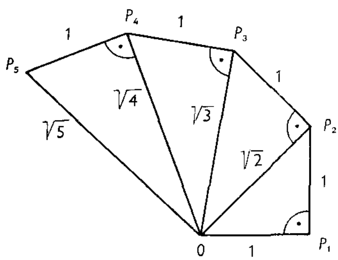
Основу математического анализа составляет дифференциальное и интегральное исчисление для вещественных функций одной вещественной переменной. Большая часть других разделов математического анализа появилась в результате дальнейшего развития дифференциального и интегрального исчисления и обобщения встречающихся в нем понятий, основными из которых являются понятия *вещественного числа*, *функции* и *предела*. С ними мы и ознакомимся прежде всего.

ВЕЩЕСТВЕННЫЕ ЧИСЛА

Рациональные числа. Число называется *рациональным*, если оно представимо в виде отношения двух целых чисел. Арифметические действия над рациональными числами известны из курса средней школы. Геометрически их можно изобразить следующим образом: выберем на горизонтальной прямой две произвольные точки, которые будем считать точками 0 и 1. Расстояние между ними примем за единицу расстояния, а направление от 0 к 1 назовем положительным направлением прямой. Точку 1 принято выбирать справа от точки 0 (см. ри-



Изображение рациональных чисел точками на числовой оси сунок), поэтому для наглядности вместо движения по прямой в положительном (отрицательном) направлении часто говорят о движении вправо (влево) от начальной точки 0. Согласно этому, числу $\frac{5}{3}$, например, соответствует точка на прямой, расположенная на расстоянии в $\frac{5}{3}$ справа от точки 0, а числу $-\frac{2}{3}$ — точка, расположенная на расстоянии в $\frac{2}{3}$ слева от нуля. Очевидно, что каждому рациональному числу соответствует одна определенная точка на построенной таким образом числовой оси. При этом множеству рациональных чисел ставится в соответствие бесконечное множество точек числовой оси. Нетрудно убедиться также и в том, что любой отрезок оси содержит рациональную точку (т. е. точку, соответствующую некоторому рациональному числу), иначе говоря, рациональные точки плотно расположены на числовой оси. Естественно, однако, поставить вопрос, всякой ли точке P числовой оси соответствует некоторое рациональное число, т. е. всякий ли отрезок $0P$ имеет рациональную длину. Ответ оказывается отрицательным. В самом деле, по теореме Пифагора очевидно, что $0P_2 = \sqrt{2}$, $0P_3 = \sqrt{3}$, $0P_4 = \sqrt{4} = 2$, $0P_5 = \sqrt{5}$, ... и т. д. Возводя в квадрат положительные целые числа, получаем



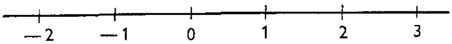
Не всякий отрезок имеет рациональную длину

$$1,\ 4,\ 9,\ 16\ \ldots$$

Квадрат дроби не может быть целым числом, т. к. из равенства

$$\frac{p^2}{q^2} = n$$

(где p, q, n — целые числа, причем дробь $\frac{p}{q}$ мы вправе считать несократимой, т. е. p и q лишенными общих множителей) следует, что дробь $\frac{p}{q}$ сократима. Но это невозможно, т. к. тогда и дробь $\frac{p^2}{q^2}$ была бы сократимой. Следовательно, если n не является квадратом некоторого целого числа, то уравнение $x^2 = n$ не имеет рационального решения. Поэтому, если отрезки $0P_2$, $0P_3$, $0P_5$ и т. д. отложить на числовой оси из начальной точки, ни один из их концов не совпадает с рациональной точкой. Иначе говоря, рациональные точки не заполняют числовую ось сплошь, остаются еще «пробелы». Уже применение теоремы Пифагора позволяет указать бесконечно много таких «пробелов», однако, значительно больше дает нам использование понятий теории множеств. Множество рациональных чисел (точек) счетно, множество же «пробелов», т. е. точек числовой оси, которым пока ничто не поставлено в соответствие, не является счетным.



Замкнутые интервалы на числовой оси

Вещественные числа. Введение вещественных чисел было обусловлено требованием поставить в соответствие каждой точке числовой оси некоторое число или, что то же самое, необходимостью численно выражать длину любого отрезка. Ниже приводится естественный способ введения этих «новых» чисел.

Пусть P — произвольная точка числовой оси (например, положительной полуоси). С помощью несложной процедуры точке P можно поставить в соответствие некоторую десятичную дробь.

Заметим, что целочисленные точки разбивают всю числовую ось на замкнутые интервалы. При этом интервал с концами a и b называется замкнутым, если точки a и b включены в его состав, и открытым — в противном случае (обозначают их, соответственно, через $[a, b]$ и (a, b)).

Если P не является точкой разбиения, она попадает в один интервал, в противном же случае из двух интервалов, общей границей которых она является, условимся выбирать правый. Назовем полученные единичные интервалы разбиения интервалами 0-го порядка. Левый конец каждого интервала будем считать его индексом. $[0, 1]$, например, является в этих обозначениях интервалом 0, $[1, 2]$ — интервалом 1, $[20, 21]$ — интервалом 20. Таким образом, точке P соответствует один определенный интервал разбиения 0-го порядка. Индекс этого интервала даст нам целую часть десятичного представления P.

В дальнейшем будем считать, что точка P попала в интервал $[0, 1]$, любой другой случай рассматривается аналогично.

Каждый интервал 0-го порядка разобьем на 10 равных интервалов 1-го порядка. Интервал $[0, 1]$, например, разбивается при этом на интервалы

$$[0;\ 0,1],\ [0,1;\ 0,2],\ [0,2;\ 0,3],\ \ldots,\ [0,9;\ 1].$$

Полученным интервалам 1-го порядка, двигаясь слева направо, присвоим индексы 0, 1, 2, 3, 4, 5, 6, 7, 8, 9. Аналогично случаю целочисленного разбиения точке P можно поставить в соответствие один из интервалов 1-го порядка, индекс которого даст нам первый десятичный знак представления P в виде десятичной дроби.

Следующим шагом является разбиение каждого интервала 1-го порядка на 10 равных интервалов 2-го порядка. Внутри каждого интервала 1-го порядка, двигаясь слева направо, присвоим интервалам 2-го порядка индексы 0, 1, 2, 3, 4, 5, 6, 7, 8, 9. Индекс того интервала 2-го порядка, который аналогично изложенному выше ставится в соответствие точке P, даст нам второй десятичный знак искомой десятичной дроби. Продолжая этот процесс дальше, можно получить сколь угодно большое число десятичных знаков. Таким образом, каждой точке P числовой оси соответствует одна определенная бесконечная десятичная дробь, причем разным точкам соответствуют разные дроби. В результате такой процедуры может появиться любая бесконечная десятичная дробь, за исключением такой, которая, начиная с некоторого места, состоит сплошь из девяток. Рациональным точкам этот процесс ставит в соответствие те же самые десятичные дроби, которые мы получили бы делением, преобразуя обыкновенные дроби в десятичные. К ним относятся конечные (которые, начиная с некоторого места, состоят только из нулей) и периодические десятичные дроби. Остальные десятичные дроби называются *иррациональными числами*.

Таким образом, мы подошли к определению вещественных чисел: так наэываются все десятичные дроби. Одновременно каждой точке числовой оси поставлено в соответствие некоторое число. Однако правильней было бы назвать десятичные дроби числами лишь в том случае, если с ними можно производить вычисления, т. е. когда определены арифметические действия над ними и установлены важные свойства, которыми эти действия обладают. Как мы увидим, вещественные числа удовлетворяют этим требованиям.

ОСНОВНЫЕ СВОЙСТВА ВЕЩЕСТВЕННЫХ ЧИСЕЛ

I. *Вещественные числа образуют поле*. Этим свойством обладают не только вещественные числа. Поля образуют также рациональные и комплексные числа (см. стр. 52). Из аксиом поля можно вывести обычные свойства действий над числами.

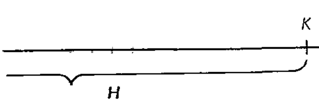
II. *Поле вещественных чисел упорядочено*. Наглядной иллюстрацией этому может служить числовая ось: $a < b$ означает, что a лежит на оси слева от b. Если a и b записаны в виде десятичных дробей, то $a < b$ означает, что первый из десятичных знаков, которые у двух дробей не одинаковы, у a меньше, чем у b.

Однако, упорядоченность числового поля означает не только то, что его элементы расположены в определенном порядке. Этот порядок должен еще согласовываться с действиями над числами, точнее, должны выполняться следующие два условия:

1. если $a < b$, то $a + c < b + c$,
2. если $a < b$ и $c > 0$, то $a \cdot c < b \cdot c$.

Поле рациональных чисел также является упорядоченным. Из свойств упорядоченности вытекают правила операций над равенствами и над абсолютной величиной.

III. *Поле вещественных чисел вполне упорядочено*. Для того, чтобы уяснить это понятие, введем несколько других понятий, которые нам понадобятся и в дальнейшем. Множество H чисел (точек числовой оси) называется *ограниченным* сверху (справа), если найдется такое число (точка) K, что ни один из элементов множества H не превышает K (ни одна из точек множества H не лежит справа от K), т. е. из $x \in H$ следует, что



Верхняя граница

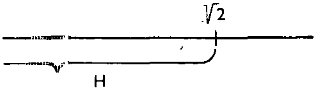
$x \leq K$. Число K называется в этом случае *верхней границей* множества H. Если множество H имеет верхнюю границу, то одновременно оно имеет бесконечно много верхних границ, т. к. любое число, большее K, очевидно, также будет верхней границей. Множество натуральных чисел, например, не ограничено сверху, т. к. каким бы мы ни выбрали число K, найдется целочисленная точка, лежащая справа от него.

Упорядоченное поле называется вполне упорядоченным, если всякое ограниченное сверху множество H имеет наименьшую верхнюю границу, т. е. такую верхнюю границу K_1, что любое число, меньшее, чем K_1, уже не является его верхней границей. Эта наименьшая из всех верхних граний называется *точной верхней границей* множества H. Для ее обозначения используют символ

$$\sup H$$

(по-латыни: supremum — наивысшее).

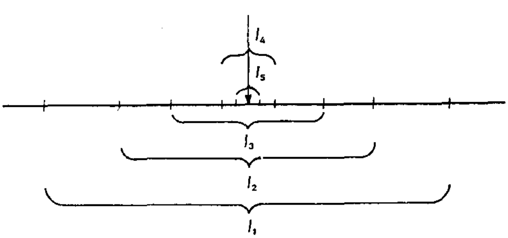
Поле рациональных чисел упорядочено, но не вполне упорядочено. В самом деле, пусть H — множество рациональных чисел, меньших $\sqrt{2}$. Очевидно, что H ограничено сверху, но не имеет точной верхней границы.



Поле вещественных чисел упорядочено, но не вполне упорядочено

Перечисленные свойства *вполне упорядоченного поля* целиком и полностью характеризуют и однозначно определяют множество вещественных чисел. Это означает, что если помимо множества бесконечных десятичных дробей мы сможем указать некоторое другое множество, которое, при соответствующем задании операций и упорядочении, также образует вполне упорядоченное поле, то это новое множество по существу не будет отличаться от множества десятичных дробей. Другими словами, мы будем иметь две различные модели или реализации одной и той же структуры. Так, например, рациональные числа могут быть заданы в виде десятичных дробей, но можно задать их и в виде обыкновенных дробей. При этом между полученным таким образом множеством десятичных дробей и множеством обыкновенных дробей может быть установлено такое взаимнооднозначное соответствие, при котором сохраняются сумма, произведение и порядок элементов. Это означает, что любые два упорядоченных поля изоморфны (см. стр. 148).

При определении полной упорядоченности полей часто не требуют существования точной верхней границы для любого ограниченного множества, заменяя это требование постулированием одного из следующих двух принципов:



Принцип вложенных интервалов

1. *Принцип вложенных интервалов*. Пусть $I_1, I_2, \ldots, I_n, \ldots$ — такая последовательность замкнутых интервалов, что каждый последующий содержится в предыдущем, причем длины интервалов стремятся к 0 с воз-

растанием n (точное определение этого будет дано ниже). Тогда существует единственная точка, принадлежащая всем интервалам (существует единственное число, которое ни при каком n не меньше левого и не больше правого конца интервала I_n).

2. *Принцип сечений.* Если область вещественных чисел разбить на два не пустые множества A и B так, чтобы каждый элемент «нижнего класса» A был меньше любого элемента «верхнего класса» B, то существует либо наибольший элемент в классе A, либо наименьший — в классе B.

Арифметические действия над вещественными числами. Ввиду обширности материала мы лишь кратко остановимся на определении арифметических действий над вещественными числами (бесконечными десятичными дробями). Пусть x — некоторая бесконечная десятичная дробь. Назовем n-ым десятичным приближением x снизу конечную десятичную дробь, состоящую из целой части и первых n десятичных знаков числа x. Например, первым приближением снизу числа $\sqrt{2}$ является 1,4; вторым приближением снизу числа $\frac{2}{3}$ является 0,66. Будем обозначать n-ое приближение числа x снизу через A_n. n-ым приближением x сверху назовем дробь

$$F_n = A_n + \frac{1}{10^n}.$$

Нетрудно видеть, что система неравенств

$$A_n \leq x \leq F_n, \qquad n = 1, 2, 3, \ldots$$

определяет последовательность вложенных интервалов, единственной общей точкой которых является x. Если теперь через $A_1, A_2, \ldots$ обозначить приближения снизу, а через $F_1, F_2, \ldots$ — приближения сверху некоторого числа x', то сумму $x + x'$ можно определить как единственную общую точку вложенных интервалов

$$[A_n + A'_n;\ F_n + F'_n], \qquad n = 1, 2, 3, \ldots$$

Аналогичным образом можно определить и произведение. Основываясь на свойствах рациональных чисел, с помощью принципа вложенных интервалов нетрудно проверить выполнение аксиом поля и правил действий над вещественными числами, однако, нередко это сопряжено с трудоемкими вычислениями.

Аналогично можно получить, например, что корень n-ой степени из положительного вещественного числа есть некоторое положительное вещественное число. В поле рациональных чисел это не выполняется.

УПОРЯДОЧЕНИЕ. НЕРАВЕНСТВА. АБСОЛЮТНАЯ ВЕЛИЧИНА

Свойства и правила операций над неравенствами и абсолютной величиной опираются на свойства упорядоченности. Можно вывести правила, применение которых — аналогично правилам операций над уравнениями — позволяет решать неравенства. Перечислим некоторые из них:

1. К обеим частям неравенства можно прибавить одно и то же число.
2. Обе части неравенства можно умножить на одно и то же число.
3. После умножения обеих частей неравенства на одно и то же отрицательное число левую и правую части неравенства надо поменять местами (или вместо $<$ написать $>$).
4. Если a и b одного знака и $a < b$, то $\frac{1}{b} < \frac{1}{a}$.
5. Неравенства можно складывать: если $a < b$ и $c < d$, то $a + c < b + d$.
6. Если $a < b$ и $0 < c < d$, то $a \cdot c < b \cdot d$.

Пример: Решим неравенство $2x + 5 < 3$.

Решение: к обеим частям неравенства $2x + 5 < 3$ прибавим -5, обе части неравенства $2x < -2$ разделим на 2, получим: $x < -1$,

т. е. решениями неравенства являются все числа, меньшие -1.

Абсолютная величина. Под абсолютной величиной числа x понимается оно само, если $x \geqq 0$, и симметричное ему число, $-x$, если $x < 0$. Абсолютную величину x можно определить формулой:

$$|x| = x, \quad \text{если} \quad x \geqq 0,$$

$$|x| = -x, \quad \text{если} \quad x < 0.$$

Абсолютная величина числа равна длине отрезка числовой оси, отмеренного из начальной точки до точки, соответствующей этому числу. Основные свойства операций над абсолютными величинами:

$$|x + y| \leqq |x| + |y|;$$

$$|x - y| \geqq ||x| - |y||;$$

$$|x \cdot y| = |x| \cdot |y|;$$

$$|x| \, |y| = |x| \, |y|.$$

Расстояние между точками a и b на числовой оси задается формулой

$$|a - b|$$

(абсолютная величина их разности).

Заметим, что $d(a, b) = |b - a|$ удовлетворяет трем основным свойствам расстояния, очевидным из наглядных соображений:

1. $d(a, b) = 0$ тогда и только тогда, когда $a = b$.
2. $d(a, b) = d(b, a)$ (свойство симметрии).
3. $d(a, b) \leqq d(a, c) + d(c, b)$ (неравенство треугольника).

Некоторые важные неравенства. С помощью свойств упорядочения и правил действий над числами можно вывести немало важных неравенств. Приведем некоторые из них:

1. *Сравнение среднего арифметического и среднего геометрического.* Среднее геометрическое неотрицательных чисел не превышает их среднего арифметического, т. е.

$$\sqrt{a \cdot b} \leqq \frac{a+b}{2}; \sqrt[3]{a \cdot b \cdot c} \leqq \frac{a+b+c}{3}; \quad (a \geqq 0,\ b \geqq 0,\ c \geqq 0).$$

Вообще

$$\sqrt[n]{a_1 a_2 \ldots a_n} \leqq \frac{a_1 + a_2 + \ldots + a_n}{n} \qquad (a_1, a_2, \ldots, a_n \geqq 0)$$

2. *Неравенство Коши:*

$$(a_1 b_1 + a_2 b_2 + \ldots + a_n b_n)^2 \leqq (a_1^2 + a_2^2 + \ldots + a_n^2) \cdot (b_1^2 + b_2^2 + \ldots + b_n^2) .$$

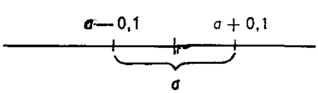
3. *Неравенство Бернулли:* если $x > -1$, то $(1 + x)^n \geqq 1 + nx$ (где n — натуральное число).

Примеры: 1. Каким неравенством можно задать открытый интервал с центром в точке a, длина которого равна 0,2?

Интервал состоит из точек x, удаленных от точки a меньше, чем на 0,1, т. е. удовлетворяющих следующему неравенству:

$$|x - a| < 0{,}1.$$

2. Пусть a и b — произвольные вещественные числа. Большее из них обозначают через $\max(a, b)$, меньшее — через $\min(a, b)$. Если $a = b$, то $\max(a, b)$ и $\min(a, b)$ считаются равными самим числам.



Открытый интервал с центром в точке a, длина которого равна 0,2

Выразим и $\max(a, b)$ и $\min(a, b)$ через уже известные нам понятия.

Очевидно,

$$\max(a, b) + \min(a, b) = a + b;$$
$$\max(a, b) - \min(a, b) = |a - b|.$$

Решая систему уравнений, получаем:

$$\max(a, b) = \frac{a + b + |b - a|}{2}; \quad \min(a,b) = \frac{a + b - |b - a|}{2}.$$

3. В случае, когда $b=0$, формулы предыдущего примера принимают вид:

$$\max(a, 0) = \frac{a+|a|}{2}; \ \min(a, 0) = \frac{a-|a|}{2}.$$

Поскольку $\max(a, 0)$ равняется a тогда и только тогда, когда a не отрицательно, эту величину называют положительной частью a. Аналогично $\min(a, 0)$ называют отрицательной частью a, т. к. он равен a тогда и только тогда, когда a не является положительным. Для положительной и отрицательной частей принято использовать следующие обозначения:

$$a_+ = \frac{a+|a|}{2}; \ a_- = \frac{a-|a|}{2}.$$

ВЕЩЕСТВЕННЫЕ ЧИСЛОВЫЕ МНОЖЕСТВА

Вещественными числовыми множествами мы называем множества, элементы которых — вещественные числа. Об этих множествах часто говорят как о множествах точек на прямой, называя их элементы точками. Эта терминология, обоснованная возможностью отождествить вещественные числа с точками числовой оси, благодаря своей наглядности позволяет проще уяснить многие понятия и теоремы. Так, например, вместо $a < b$ можно сказать, что a лежит слева от b, вместо $a \leqq b$ — что a лежит не правее b, а вместо $a < b < c$, что b лежит между a и c. Ознакомимся вкратце с основными понятиями теории точечных множеств (множеств точек на прямой).

Основные понятия теории точечных множеств. Множество H точек прямой называется ограниченным сверху, если существует такое число K, что для любого элемента x множества H $x \leqq K$, т. е. ни один из элементов H не лежит справа от K.

Наименьшая из верхних границ множества H называется его точной верхней границей ($\sup H$). Если множество H состоит из конечного числа элементов, его точной верхней границей является наибольший из элементов. Естественно, это остается верным и в случае бесконечного множества H, имеющего максимальный элемент. Однако, даже ограниченное сверху бесконечное множество не всегда имеет максимальный элемент. Например, у множества отрицательных чисел нет максимального элемента. Тем не менее и в этом случае, используя уже знакомый нам процесс отыскания десятичных приближений, с помощью которого мы определили вещественные числа как десятичные дроби, нетрудно убедиться, что если множество H ограничено сверху, то оно имеет точную верхнюю границу. В самом деле, рассмотрим сначала отрезки целочисленного разбиения (для простоты изложения предположим, что элементы H — положительные числа):

$$[0;\ 1],\ [1;\ 2],\ [2;\ 3], \ldots$$

Их индексы в порядке возрастания: $0, 1, 2, \ldots$ Выберем самый правый из отрезков, содержащих элементы H. Его индекс даст нам целую часть

$\sup H$. Пусть, например, он равен 2. Тогда рассмотрим отрезки

$$[2;\ 2{,}1],\ [2{,}1;\ 2{,}2],\ \ldots,\ [2{,}9;\ 3],$$

и выберем опять самый правый из содержащих элементы H. Если, например, это отрезок $[2{,}3;\ 2{,}4]$, то первый десятичный знак $\sup H$ есть 3. Продолжая этот процесс дальше, получим следующие десятичные знаки $\sup H$.

Аналогично понятию точной верхней границы для множества H, ограниченного снизу (т. е. слева), вводится понятие точной нижней границы как наибольшей из всех нижних границ множества H. Для ее обозначения используется символ

$$\inf H$$

(от латинского infimum — самое низкое).

Множество точек, удаленных от a меньше, чем на ε, т. е. удовлетворяющих неравенству

$$|x-a|<\varepsilon.$$

называется ε — окрестностью точки a числовой оси (вещественного числа a).

Например, 0,001-окрестностью для числа 3 является открытый интервал $(2{,}999;\ 3{,}001)$.

Открытое множество. Открытым называется множество, содержащее каждую свою точку вместе с некоторой ее окрестностью (ε-окрестностью).

Примеры:

1. Всякий открытый интервал есть открытое множество.

2. Замкнутый интервал не является открытым множеством, т. к. не содержит ни одной окрестности своих крайних точек.

Из определения открытого множества очевидно, что

а) объединение открытых множеств есть открытое множество;

б) пересечение конечного числа открытых множеств есть открытое множество;

в) пустое множество и вся числовая прямая есть открытые множества.

Точка a называется предельной точкой множества H, если любая ее окрестность содержит бесконечно много элементов H.

Примеры:

1. Точка 0 является предельной точкой множества $\left\{1, \frac{1}{2}, \frac{1}{3}, \ldots\right\}$, т. к. в любой окрестности нуля содержится бесконечно много чисел вида $\frac{1}{n}$; других предельных точек у этого множества нет.

2. Предельными точками открытого интервала являются все его точки а также две его крайние точки.

3. Предельными точками множества рациональных чисел являются все точки числовой оси, т. к. любая окрестность произвольной точки оси содержит бесконечно много рациональных чисел.

4. Конечное множество не имеет предельных точек.

5. Множество целых чисел не имеет предельных точек.

Множество в последнем примере имеет бесконечное число элементов, но все же не имеет предельной точки. В этом нетрудно убедиться, поскольку в любой интервал числовой оси попадает конечное число элементов этого множества. Если же существует такой интервал, в который попадает бесконечно много элементов множества H, то H имеет предельную точку. Это утверждает следующая очень важная

теорема Больцано—Вейерштрасса: всякое бесконечное ограниченное множество вещественных чисел имеет предельную точку.

Этот результат также нетрудно получить с помощью знакомого нам процесса последовательных десятичных приближений. Если, например, в интервале $[0;\ 1]$ содержится бесконечно много точек из H, то по крайней мере в один из интервалов $[0;\ 0{,}1], \ldots, [0{,}9;\ 1]$ попадает бесконечно много точек из H. Пусть, например, это будет интервал $[0{,}2;\ 0{,}3]$. В этом случае первым десятичным приближением предельной точки будет 0,2. Продолжая процесс десятичных приближений, мы получим некоторую десятичную дробь, всякая окрестность которой содержит бесконечно много элементов H, т. е. получим предельную точку множества H.

Замкнутое множество. Множество H называется замкнутым, если оно содержит все свои предельные точки.

Примеры:

1. Всякий замкнутый интервал есть замкнутое множество.

2. Множество $\left\{1, \frac{1}{2}, \frac{1}{3}, \ldots, \frac{1}{n}, \ldots, 0\right\}$ замкнуто, т. к. содержит единственную свою предельную точку.

3. Множество всех рациональных чисел не является замкнутым.

4. Открытый интервал $(0;\ 1)$ не является замкнутым множеством.

5. Множество $\left\{1, \frac{1}{2}, \frac{1}{3}, \ldots, \frac{1}{n}, \ldots\right\}$ не является замкнутым. Из определения замкнутого множества следуют, очевидно, свойства:

а) пересечение замкнутых множеств замкнуто;

б) объединение конечного числа замкнутых множеств замкнуто;

в) множество всех вещественных чисел замкнуто;

г) пустое множество замкнуто.

Открытыми и замкнутыми множествами, однако, далеко не исчерпываются все множества, т. к. вся числовая ось, например, является одновременно и открытым, и замкнутым множеством, а полуоткрытый интервал $[1;\ 2]$ не является ни открытым, ни замкнутым. Связь между открытыми и замкнутыми множествами следующая: дополнение (см. «Теория множеств») всякого открытого множества замкнуто, дополнение всякого замкнутого множества открыто.

Выше было сказано, что рациональные точки *плотно* лежат на числовой оси. Это означает, что в промежутке между любыми двумя точками числовой оси найдется рациональная точка. Вообще некоторое множество H называется *плотным* (на числовой оси), если всякий открытый интервал числовой оси содержит точку из H. Например, множество иррациональных точек плотно.

Множество H называется связным, если вместе с любыми двумя своими точками оно содержит и интервал между ними. Другими словами, H — связное множество, если любые две его точки могут быть соединены отрезком внутри множества H.

Нетрудно видеть, что связными множествами на прямой являются следующие: пустое множество, вся числовая прямая, все множества, состоящие из одной точки, всевозможные интервалы (открытые, замкнутые и полуоткрытые) и все полупрямые. Полупрямые принято обозначать следующим образом:

$$(-\infty;\ a],\ (-\infty;\ a),\ [a;\ \infty),\ (a;\ \infty).$$

Здесь $[a;\ \infty)$ обозначает множество точек числовой оси, которые лежат не левее a; $(a;\ \infty)$ — множество точек, лежащих справа от a. Смысл двух других обозначений аналогичен.

ПОНЯТИЕ ФУНКЦИИ И СПОСОБЫ ЕЕ ЗАДАНИЯ

Если между двумя переменными величинами существует зависимость, при которой значение одной из них однозначно определяет значение другой, то говорят, что вторая величина есть функция первой. При этом первую величину называют независимой, а вторую — зависимой переменной. Зависимость между ними называется функциональной. Вообще мы говорим о функциональной зависимости, если каждой точке x некоторого множества H с помощью некоторого правила f однозначно ставится в соответствие некоторое число $y = f(x)$. Множество H в этом случае называется областью определения, а множество K значений y, соответствующих точкам из H — областью значений функции.

Областью определения функции $y = \sin x$, например, является вся числовая ось, а областью значений — интервал $[-1;\ 1]$.

Считая, что значение независимой переменной x есть некоторая точка P числовой оси, а соответствующее значение функции $y = f(x)$ — точка P' другой числовой оси, функцию называют отображением, а точку P' — образом точки P.

Примеры:

1. В предположении, что температура некоторой массы идеального газа сохраняется неизменной, объем v и давление p этой массы газа под-

чиняются закону Бойля—Мариотта:

$$pv = C\,,$$

где C — некоторая константа. Отсюда

$$p = \frac{C}{v},$$

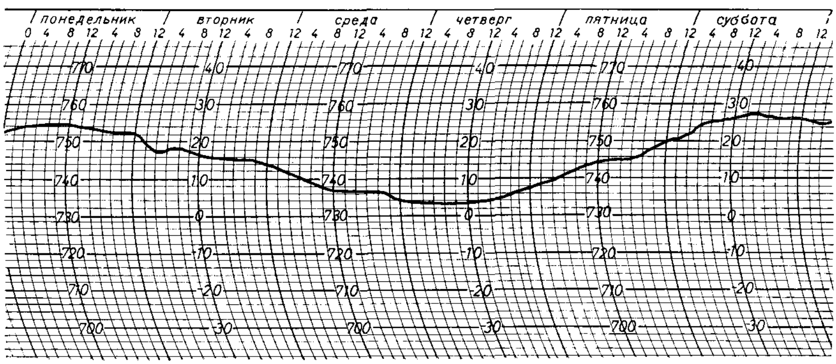
т. е. величина объема однозначно определяет величину давления, иными словами, давление есть функция объема.

2. Площадь круга $t = \pi r^2$, т. е. площадь t есть функция радиуса r.

3. Пусть $y = p(n)$ определяет n-ое простое число, тогда

$$p(1) = 2,\ p(2) = 3,\ p(3) = 5,\ p(4) = 7 \text{ и т. д.}$$

4. Пусть $y = \sqrt{1 - x^2}$. Значение y определяется как функция от x.



Давление воздуха как функция времени

5. Пишущее устройство барометра-анероида рисует кривую на движущейся бумажной ленте. Эта кривая иллюстрирует зависимость между давлением воздуха и временем. С ее помощью в любое время дня можно узнать величину давления воздуха (как функцию времени).

6. Хорошо известными примерами функций являются показательная, логарифмическая и тригонометрические функции (см. главу «Алгебра»).

Как видно из приведенных примеров, функцию можно задавать

1. аналитически, т. е. математической формулой, дающей возможность вычислять значения функции;

2. графически;

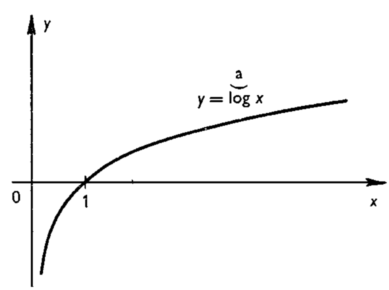


График функции $y = \log_a x$

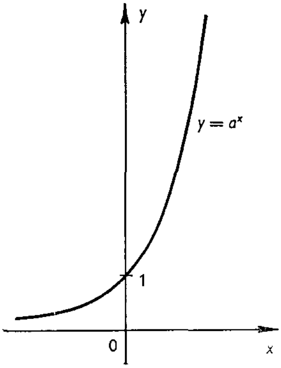


График функций $y = a^x$

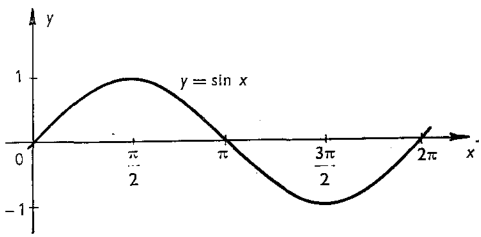


График функции $y = \sin x$

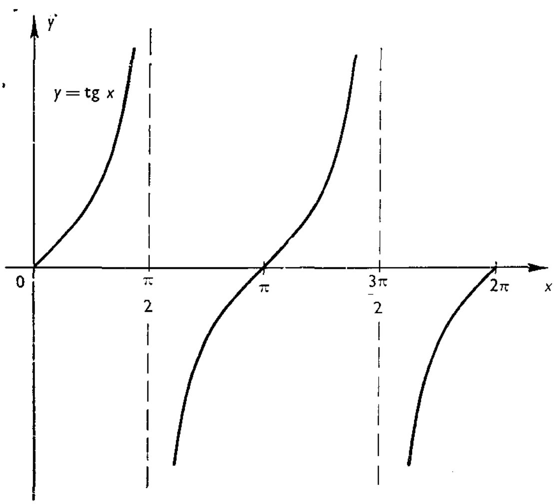


График функции $y = \operatorname{tg} x$

3. при помощи таблицы;

4. при помощи словесного описания.

Графический способ задания позволяет наглядно проследить ход изменения функций, поэтому к его помощи целесообразно прибегать, насколько это возможно, и в тех случаях, когда функции заданы другим способом.

Так, например, ход изменения логарифмической функции, заданной таблицей логарифмов, становится наглядным и легко запоминающимся, если нарисовать логарифмическую кривую. То же самое можно сказать относительно показательной и тригонометрических функций.

Пусть на плоскости задана система координат. Графиком функции называется множество точек $(x; y)$, где x есть элемент области определения функции, а y — соответствующее ему значение функции.

При таком определении графика не вызывает затруднений и случай, когда функцию нельзя изобразить. Такова, например, функция Дирихле:

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

НЕКОТОРЫЕ ВАЖНЫЕ ТИПЫ ФУНКЦИЙ

Функция $y = f(x)$ называется *ограниченной*, если область ее значений есть ограниченное множество. Например, функция $y = \sin x$ ограничена на всей прямой, а функция $y = \operatorname{tg} x$ не ограничена на всей прямой, но ограничена на отрезке $\left[-\frac{\pi}{4}; \frac{\pi}{4}\right]$.

Функция $y = f(x)$ называется *монотонно возрастающей*, (в строгом смысле) на множестве H, если из $x_1 < x_2$ следует, что $f(x_1) < f(x_2)$, и *монотонно убывающей*, (в строгом смысле), если $f(x_1) > f(x_2)$ при $x_1 < x_2$ (для любых x_1 и x_2 из H). Функция $y = x^2$, например, монотонно убывает на полупрямой $(-\infty; 0]$ и монотонно возрастает на полупрямой $[0; \infty)$. Если же из $x_1 < x_2$ следует лишь $f(x_1) \leqq f(x_2)[f(x_1) \geqq f(x_2)]$, то функция называется *неубывающей (невозрастающей)*. Постоянная функция, например, является одновременно и неубывающей, и невозрастающей.

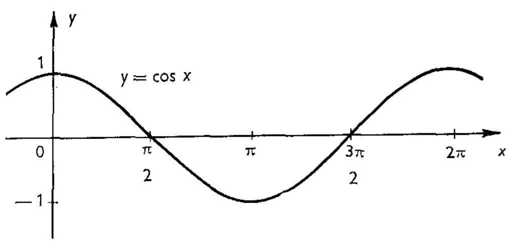


График функции $y = \cos x$

Функция $y = f(x)$ называется *четной*, если

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

и нечетной, если $f(x) = -f(-x)$. График четной функции симметричен относительно оси y, а график нечетной — относительно начала координат. Четными, например, являются функции $y = x^n$, при четном n, и $y = \cos x$, а нечетными — функции $y = x^n$ при n нечетном и $y = \sin x$. Сами названия четной и нечетной функций происходят от свойства функции $y = x^n$ быть таковой при четном, соответственно нечетном n.

Функция $y = f(x)$ называется *периодической*, если $f(x + a) = f(x)$ для всех x при каком-нибудь постоянном $a > 0$. Наименьшее среди a, для которых это выполняется, называется *периодом* функции $f(x)$. Функция $y = \sin x$, например, периодическая с периодом 2π, функция $y = \operatorname{tg} x$ также периодическая, но с периодом π.

ПОНЯТИЕ ОБРАТНОЙ ФУНКЦИИ

Если соответствие, устанавливаемое функцией $f(x)$ между областью ее определения H и областью значений K таково, что каждый элемент области K является образом ровно одного элемента области H, то, рассматривая обратное отображение, мы также получим функцию. Действительно, в этом случае каждый элемент множества K однозначно определяет тот элемент множества H, образом которого он является при отображении f. Получаемую таким образом функцию называют функцией, обратной для функции f, и часто обозначают через f^{-1}. Например, для функции $y = \lg x$

обратной является функция $y = 10^x$. Для функции $y = x^2$, если ее областью определения считать всю числовую ось, нет однозначно определенной обратной функции. Причина заключается в том, что, например, значение $y = 4$ функция принимает в двух точках: $x = 2$, и $x = -2$, поэтому в точке 4 обратную функцию нельзя определить однозначно. Если же в качестве области определения функции $y = x^2$ рассмотреть лишь полупрямую $[0;\ \infty)$, то однозначной обратной функцией для нее будет $y = \sqrt{x}$.

Для функций, заданных аналитически, обратную функцию можно получить, выразив x через y. После этого, следуя традиции обозначать независимую переменную буквой x, а зависимую — буквой y, условимся менять x и y местами.

Примеры:

1. $y = 2\lg x - 5$, отсюда

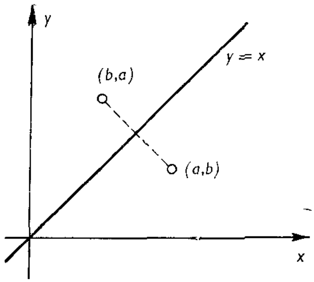
$$\lg x = \frac{y+5}{2}, \quad x = 10^{\frac{y+5}{2}},$$

т. е. обратной функцией является $y = 10^{\frac{x+5}{2}}$.

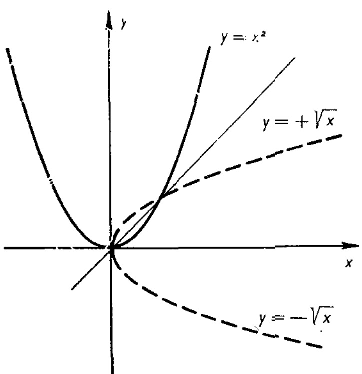
2. $y = \dfrac{2x+3}{x-1}$, отсюда

$$x = \frac{y+3}{y-2},$$

т. е. обратная функция имеет вид: $y = \dfrac{x+3}{x-2}$.



Зеркальное отображение точки (a, b) относительно прямой $y = x$

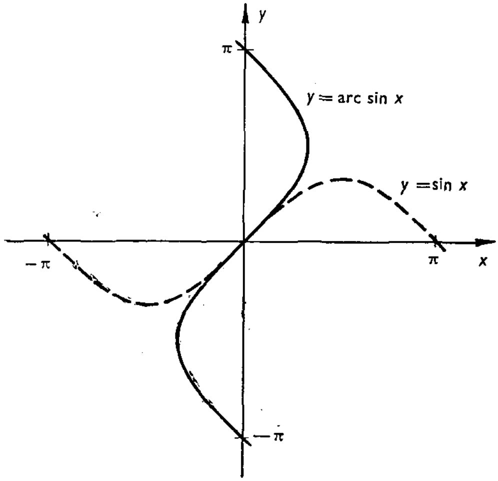


Графики функции $y = x^2$ и обратной ей функции $y = +\sqrt{x}$ и $y = -\sqrt{x}$

Если функция задана графически, то условие, что она должна принимать каждое свое значение не более, чем в одной точке, т. е. условие существования однозначной обратной функции, сводится к требованию, чтобы всякая прямая, параллельная оси x, пересекала график функции не более, чем в одной точке.

График обратной функции получается из графика самой функции зеркальным отражением относительно прямой $y = x$. Это основано на том, что зеркальным отражением произвольной точки (a, b) относительно прямой $y = x$ является точка (b, a).

Понятие непрерывности функции мы сможем определить строго лишь после ознакомления с понятием предела. Од-



Графики функции $y = \sin x$ и обратной ей функции $y = \arcsin x$

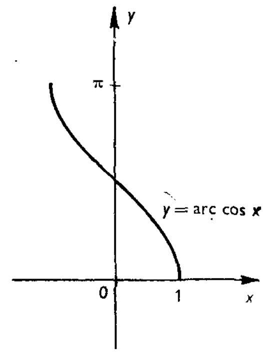


График функции $y = \arccos x$

нако, наглядное представление о непрерывности можно получить, если, следуя историческому развитию этого понятия, назвать непрерывными те функции, графики которых могут быть нарисованы в виде непрерывной (сплошной) линии. Ниже мы увидим, что такая трактовка слишком узка, но в первом приближении вполне достаточна. Можно доказать, что для функции, непрерывной в некотором интервале, однозначная обратная функция существует в том случае, если сама функция строго монотонна (возрастает или убывает). Если же непрерывная функция не является сторого монотонной во всей своей области определения, то, если это возможно, область определения разбивают на ин-

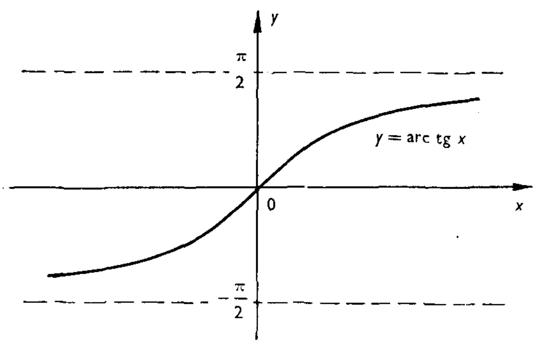


График функции $y = \operatorname{arc\,tg} x$

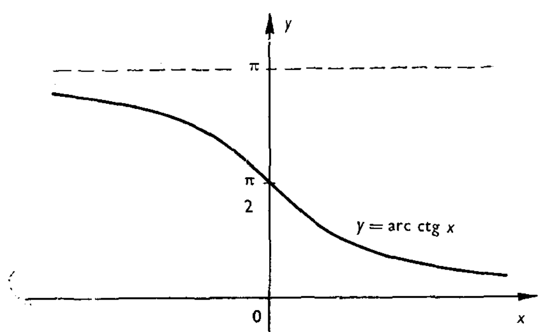


График функции $y = \operatorname{arc\,ctg} x$

тервалы, в которых функция строго монотонна. В каждом таком интервале монотонности для функции, очевидно, уже существует однозначная обратная функция.

Примеры:

1. Для функции $y = x^2$, рассматриваемой на полупрямой $(-\infty;\ 0]$, обратной является функция $y = -\sqrt{x}$. Для ветви функции $y = x^2$, определенной в интервале $(0;\ \infty)$, обратной функцией является $y = \sqrt{x}$.

2. Для определения однозначной «ветви» функции, обратной функции $y = \sin x$, выберем один из интервалов монотонности синуса. Обычно принято рассматривать интервал $\left[-\frac{\pi}{2};\frac{\pi}{2}\right]$, в котором обратной для $y = \sin x$ является функция

$$y = \arcsin x\,,$$

изображенная на рисунке (стр. 397).

Аналогичным образом можно определить однозначные «ветви» функций, обратных функциям $y = \cos x$, $y = \operatorname{tg} x$ и $y = \operatorname{ctg} x$.

ПОНЯТИЕ СЛОЖНОЙ ФУНКЦИИ

Если функция $y = f(u)$ отображает область определения K в область значений L, а функция $u = g(x)$ отображает свою область определения H в область значений K_1, которая является частью области K, тогда функция

$$y = f(g(x))\,,$$

составленная из двух функций, отображает область определения H в область значений L.

Пример: Функция $y = \sin u$ отображает числовую ось $(-\infty;\ \infty)$ в интервал $[-1;\ 1]$; функция $u = x^2 + 1$ отображает числовую ось в полупрямую $[1;\ \infty)$. Сложная функция

$$y = \sin(x^2 + 1)\,,$$

составленная из двух функций, отображает числовую ось $(-\infty;\ \infty)$ в интервал $[-1;\ 1]$.

Разумеется, сложную функцию можно составлять и из большего числа функций, например:

$$y = \lg u;\quad u = v^2 + 2;\quad v = \sin x;$$
$$y = \lg(\sin^2 x + 2)\,.$$

ЭЛЕМЕНТАРНЫЕ ФУНКЦИИ

Если функции $f(x)$ и $g(x)$ определены на одном и том же множестве H, то, используя действия над числами, можно очевидным образом определить алгебраические операции над функциями. Под суммой двух функций f_1 и f_2 понимают такую функцию f, которая в каждой точке x области определения H принимает значение, равное сумме значений f_1 и f_2 в той же точке, т. е.

$$f(x) = f_1(x) + f_2(x) .$$

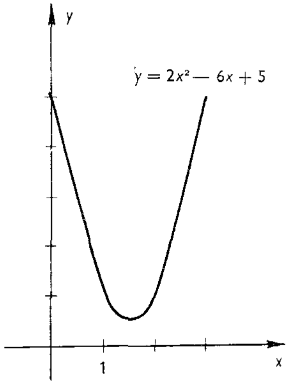


График целой рациональной функции

Аналогично определяются разность и произведение двух функций. При определении отношения двух функций область их определения необходимо сузить, исключив из нее те точки, в которых знаменатель обращается в 0. Например, областью определения функций $\sin x$ и $\cos x$ является вся числовая ось, однако, при определении функции $\operatorname{tg} x$, их отношения, следует исключить из рассмотрения точки числовой оси, в которых $\cos x$ обращается в нуль, т. е. точки вида $(2k+1)\frac{\pi}{2}$, $k = 0, \pm 1, \pm 2, \ldots$

Рассмотрим теперь множество всех тех функций, которые могут быть получены в результате выполнения конечного числа операций сложения, вычитания и умножения из следующих двух простых функций: $y = C$ (постоянная) и $y = x$. Мы получим класс так называемых целых рациональных функций (многочленов). Общий вид целой рациональной функции:

$$y = a_0x^n + a_1x^{n-1} + \ldots + a_{n-1}x + a_n .$$

Целые рациональные функции с полным правом можно назвать самыми простыми функциями, т. к. значение функции $f(x)$ при данном значении x независимой переменной можно вычислить, выполнив конечное число операций сложения, вычитания и умножения. Сравнительно несложно построить график такой функции. Так, на рисунке изображен график целой рациональной функции

$$y = 2x^2 - 6x + 5 .$$

Сумма, разность и произведение двух произвольных целых рациональных функций также является целой рациональной функцией.

Если же рассматривать множества функций, которые содержат функции $y = C$ и $y = x$ и, кроме того, вместе с любыми двумя своими элементами содержат также их сумму, разность, произведение и отношение, то как минимальное из множеств, удовлетворяющих этим требованиям, мы получим класс дробных рациональных функций. Общий вид дробной

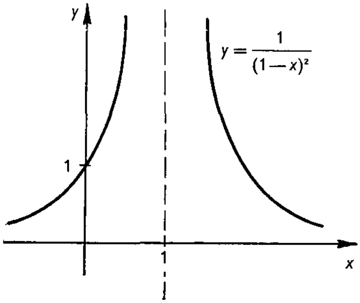


График функции $y = \dfrac{1}{(1-x)^2}$, имеющей разрыв при $x = 1$

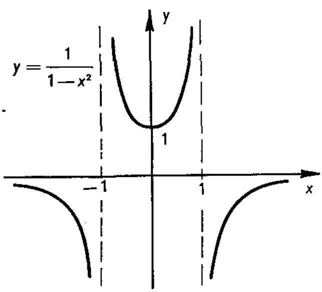


График функции $y = \dfrac{1}{1-x^2}$, имеющей разрывы при $x = 1$ и $x = -1$

рациональной функции:

$$y = \frac{a_0x^n + a_1x^{n-1} + \ldots + a_n}{b_0x^m + b_1x^{m-1} + \ldots + b_m},$$

т. е. всякая дробная рациональная функция представима в виде отношения двух целых рациональных функций. Вычисление значения дробной рациональной функции в данной точке также не представляет затруднений, поскольку требует лишь выполнения конечного числа операций сложения, вычитания, умножения и деления.

Графики дробных рациональных функций строятся, вообще говоря, сложнее, чем графики целых рациональных функций, т. к. в этом случае в точках, где знаменатель функции обращается в нуль, имеются различного рода разрывы, как это видно на рисунках, где изображены, соответственно, графики функций

$$y = \frac{1}{(1-x)^2} \quad \text{и}$$

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

Сумма, разность, произведение и отношение двух дробных рациональных функций являются дробными рациональными функциями; сложная функция, составленная из дробных рациональных функций, также оказывается дробной рациональной функцией. Однако, при отыскании функций, обратных целым или дробным рациональным, приходится, как правило, выходить за пределы этого класса функций. Например, для функции $y = x^n$ обратной является функция $y = \sqrt[n]{x}$, которая уже не принадлежит классу дробных рациональных функций.

АЛГЕБРАИЧЕСКИЕ ФУНКЦИИ

Функции, обратные дробным рациональным, принадлежат более широкому классу: множеству так называемых *алгебраических функций*. Говорят, что y является алгебраической функцией x, если зависимость между ними выражается соотношением вида

$$p_0(x)y^n + p_1(x)y^{n-1} + \ldots + p_{n-1}(x)y + p_n(x) = 0,$$

где

$$p_0(x), \quad p_1(x), \ldots, p_n(x)$$

многочлены относительно x. Кроме этого необходимо, чтобы переменная y удовлетворяла определенным требованиям, на которых мы сейчас не будем останавливаться.

Примеры алгебраических функций:

1. Всякая дробная рациональная функция является алгебраической. Действительно, если

$$y = \frac{p(x)}{q(x)}, \quad \text{то}$$

$$q(x)y - p(x) = 0 .$$

2. Корень n-ой степени из дробной рациональной функции является алгебраической функцией. В самом деле, если

$$y = \sqrt[n]{\frac{p(x)}{q(x)}}, \quad \text{то}$$

$$q(x)y^n - p(x) = 0 .$$

3. Алгебраической является всякая функция, образованная из переменной величины x и постоянных величин после применения к ним конечного числа алгебраических операций и извлечения корней. Такова, например, функция

$$y = \sqrt[3]{\frac{\sqrt{5x^2+2} + \sqrt[5]{2x^2+3}}{x^4+1}} .$$

ТРАНСЦЕНДЕНТНЫЕ ФУНКЦИИ

рансцендентной функцией называется всякая функция, не являющаяся алгебраической. Таковы, например, показательная функция, логарифмическая функция, тригонометрические и обратные им функции. Следует еще упомянуть так называемые гиперболические функции, которые часто встречаются и также являются трансцендентными. К ним относятся:

$$\operatorname{sh} x = \frac{1}{2}(e^x - e^{-x}); \quad \operatorname{ch} x = \frac{1}{2}(e^x + e^{-x});$$

$$\operatorname{tgh} x = \frac{\operatorname{sh} x}{\operatorname{ch} x}; \qquad \operatorname{ctgh} x = \frac{\operatorname{ch} x}{\operatorname{sh} x} .$$

Элементарными функциями мы называем все те функции, которые могут быть получены из дробных рациональных, показательной, логарифмической, тригонометрических и гиперболических функций, а также из функций, обратных им, с помощью конечного числа алгебраических операций, извлечения корней и образования сложных функций. Такова, например, функция

$$y = \sqrt[5]{\sin^2 x + \lg x} - \frac{2^x + 1}{\sqrt[4]{x^2 + x + 1}}.$$

Формулы, которыми задаются элементарные функции, мы называем *аналитическими выражениями*.

Очевидно, само название элементарных функций сложилось исторически. Как мы видели, рациональные функции можно с полным правом назвать самыми простыми. Однако, что касается математики, то нет никакого основания назвать элементарные трансцендентные функции (показательную, тригонометрические, гиперболические и обратные им функции) более простыми, чем многие другие трансцендентные функции. Ведь функция $y = [x]$, например, которая означает наибольшее целое число, не превышающее x, или уже упоминавшаяся функция Дирихле:

$$f(x) = \begin{cases} 0, & \text{если } x \text{ иррациональное,} \\ 1, & \text{если } x \text{ рациональное,} \end{cases}$$

являются трансцендентными функциями, но они, очевидно, ничем не сложнее элементарных трансцендентных функций. Дело просто в том, что в процессе развития математики и ее приложений элементарные трансцендентные функции появились сравнительно рано и играли важную роль, поэтому и символы, введенные для их обозначения, как, например, $\sin x$, стали хорошо известными и привычными.

ПРЕДЕЛ ЧИСЛОВОЙ ПОСЛЕДОВАТЕЛЬНОСТИ

Числовой последовательностью называют такую функцию, областью определения которой является множество положительных целых чисел.

Для обозначения последовательностей, по установившемуся соглашению, используют способ, отличающийся от известного способа обозначения функций: вместо $a(1), a(2), \ldots, a(n), \ldots$ пишут $a_1, a_2, \ldots, a_n, \ldots$, т. е. независимая переменная используется в качестве индекса. Кроме того, a_n-ое принято называть не значением в точке n, а n-ым членом последовательности. Для обозначения всей последовательности принято использовать символ $\{a_n\}$.

Рассмотрим несколько простых примеров последовательностей.

1. $$a_n = \frac{1}{n^2}.$$

Первые ее члены:

$$a_1 = 1, \quad a_2 = \frac{1}{4}, \quad a_3 = \frac{1}{9}, \ldots$$

Последовательность $a_n = \frac{1}{n^2}$

Последовательность ограничена и монотонно убывает. Изображая члены последовательности на числовой оси, видим, что они стремятся к нулю. Это означает, что в сколь угодно малую окрестность точки 0 попадают все члены последовательности, для которых n достаточно велико. Например, $\frac{1}{n^2}$ попадает в $\frac{1}{100}$-окрестность 0, если $n > 10$, т. е. $a_{11}, a_{12}, a_{13}, \ldots$ лежат ближе к нулю, чем $\frac{1}{100}$. Если рассмотреть меньшую окрестность, например, 10^{-10}-окрестность нуля, в нее попадут все a_n, для которых $n > 10^5$. Про такую последовательность говорят, что она *сходится* к своему *пределу* 0 (по-латыни: *limes*), когда n стремится к ∞, и обозначают:

$$\lim_{n\to\infty} \frac{1}{n^2} = 0.$$

Число 0 при этом называют пределом последовательности.

2. $$a_n = (-1)^n.$$

Последовательность ограничена, но не монотонна и, очевидно, не сходится ни к какому пределу, поскольку не существует числа, от которого и -1, и 1 отличались бы одновременно сколь угодно мало, например, меньше, чем на $\frac{1}{3}$. Такие (не сходящиеся) последовательности называют *расходящимися*.

3. $$a_n = \frac{2n-1}{n} = 2 - \frac{1}{n}; \quad a_1 = 1, \quad a_2 = \frac{3}{2}; \quad a_3 = \frac{5}{3}, \quad \ldots$$

Последовательность ограничена, монотонно возрастает и сходится к 2, т. к. ее члены отличаются от 2 на величину $\frac{1}{n}$, которая сколь угодно мала при достаточно большом n, т. е.

$$\lim_{n\to\infty} \frac{2n-1}{n} = 2.$$

4. $$a_n = (-1)^{n-1} \cdot \frac{1}{n}; \quad a_1 = 1, \quad a_2 = -\frac{1}{2}, \quad a_3 = \frac{1}{3}, \quad \ldots$$

Последовательность ограничена, не монотонна и сходится. Ее пределом является 0.

5. $a_n = p_n$ n-ое простое число, т. е. $a_1 = 2$, $a_2 = 3$, $a_3 = 5, \ldots$ Последовательность монотонно возрастает, не ограничена и, очевидно, не сходится ни к какому числу. Члены последовательности сколь угодно велики при достаточно большом n. Это означает, что $p_n > M$ для любого наперед заданного числа M, если n достаточно велико. Насколько велико должно быть n — зависит от числа M. Про такую последовательность говорят, что она стремится к ∞, когда n стремится к ∞ и обозначают:

$$\lim_{n \to \infty} p_n = \infty.$$

Сходящимися называются последовательности, имеющие конечный предел, поэтому последовательности, стремящиеся к ∞, мы называем расходящимися. Аналогично вводится понятие последовательности, стремящейся к $-\infty$. Такова, например, последовательность $a_n = 1 - 2n$.

6. $$a_n = (-1)^n n .$$

Последовательность не ограничена и не стремится ни к ∞, ни к $-\infty$.

ПОНЯТИЕ СХОДИМОСТИ

Основываясь на приведенных примерах, сформулируем теперь общее определение понятий сходимости и предела числовой последовательности.

Говорят, что последовательность a_n сходится к числу l, если в сколь угодно малую окрестность числа l попадают все члены последовательности, индекс которых достаточно велик, или, что то же самое, если члены последовательности отличаются от числа l меньше, чем на сколь угодно малое положительное число ε, когда n достаточно велико.

В обычных обозначениях

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

означает, что для любого $\varepsilon > 0$ существует такое целое положительное число $N(\varepsilon)$, что

$$| a_n - l | < \varepsilon \qquad \text{если} \qquad n > N(\varepsilon) .$$

Значение $N(\varepsilon)$ иногда называют порогом, соответствующим числу ε. Вместо $N(\varepsilon)$ можно писать просто N, не подчеркивая зависимости этого числа от ε. Вместо $\lim\limits_{n \to \infty} a_n = l$ иногда пишут $a_n \to l$.

Если последовательность стремится к какому-нибудь числу, то ее называют сходящейся, в противном случае — расходящейся.

Если последовательность $\{a_n\}$ такова, что для произвольно заданного числа M найдется число N такое, что

$$a_n > M, \qquad \text{если} \qquad n > N,$$

то принято считать, что $\lim\limits_{n \to \infty} a_n = \infty$.

Итак, мы познакомились с понятиями сходимости и предела числовой последовательности. Следует, однако, сказать несколько слов о том, как и для чего эти понятия используются. В самом деле, очевидно, например, что

$$\lim_{n \to \infty} \left(A + \frac{1}{n} \right) = A, \quad \lim_{n \to \infty} n^2 = \infty,$$

но до сих пор мы никак эти пределы не использовали. Приведенные выше примеры вряд ли были сколь-нибудь интереснее: их целью являлась лишь иллюстрация вновь введенных понятий.

Вспомним теперь о том факте, что площадь круга является пределом площадей правильных вписанных многоугольников, число сторон которых неограниченно возрастает. Из наглядных соображений очевидно, что речь идет о пределе последовательности, однако, подобная формулировка не представляет слишком большого интереса. Иначе выглядит то же самое утверждение, если за отправную точку принять тот факт, что площади вписанных многоугольников мы можем вычислить, в то время, как площадь круга неизвестна. В этом случае задана последовательность $\{t_n\}$ площадей правильных n-угольников, но неизвестно, сходится ли она, а если сходится, то к какому пределу. Если удастся доказать, что последовательность $\{t_n\}$ сходится, то, приняв во внимание некоторые другие факторы, связанные с понятием площади, можно определить площадь круга как величину

$$t = \lim_{n \to \infty} t_n.$$

Если n достаточно велико, то значения t можно с любой степенью точности приблизить значениями t_n.

Как видно из приведенного примера, проблема заключается чаще всего не в том, чтобы выяснить, сходится ли данная последовательность к данному пределу или нет, а в необходимости установить, является ли данная последовательность сходящейся, и, если да, то найти ее предел.

КРИТЕРИЙ СХОДИМОСТИ ДЛЯ МОНОТОННЫХ ПОСЛЕДОВАТЕЛЬНОСТЕЙ

Критерием сходимости мы называем такую теорему, применение которой позволяет установить, сходится ли данная последовательность или нет. Очень часто применяется следующая теорема:

Всякая ограниченная монотонно возрастающая числовая последовательность сходится и ее пределом является точная верхняя граница области ее значений. Аналогично, ограниченная монотонно убывающая числовая последовательность также сходится, а ее пределом является точная нижняя граница области значений.

Доказательство теоремы чрезвычайно просто. Если, например, последовательность $\{a_n\}$ монотонно возрастает, и точной верхней границей области ее значений является число l, то для всех членов последовательности $a_n < l$. С другой стороны, если $\varepsilon > 0$, то $l - \varepsilon$ уже не является точной верхней границей области значений последовательности, поэтому существует такое целое число N, что $a_N > l - \varepsilon$, но тогда, поскольку последовательность монотонна, $a_n > l - \varepsilon$ для всех $n > N$. Итак, при $n > N$ имеем:

$$l - \varepsilon < a_n \leq l,$$

т. е. члены последовательности, для которых $n > N$, попадают в ε-окрестность точки l.

Если же монотонно возрастающая последовательность $\{a_n\}$ не ограничена, то

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

Точно так же не ограниченная монотонно убывающая последовательность стремится к $-\infty$.

Примеры:

1. Последовательность $a_n = 2^n$ монотонно возрастает, но не ограничена. Это видно из того, что $2^n > n$ (доказывается применением полной индукции).

2. Последовательность $a_n = 0{,}5^n$ монотонно убывает и ограничена (все ее члены положительны), следовательно, сходится. То же самое верно для всякой последовательности $a_n = x^n$, если $0 < x < 1$.

3. Последовательность $a_n = \left(1 + \dfrac{1}{n}\right)^n$ монотонно возрастает и ограничена, ее предел обозначают буквой e. Число $e = 2{,}7182\ldots$ — иррациональное. Логарифмы по основанию e называются *натуральными* и обозначаются знаком ln (сокращение латинских слов logaritmus naturalis). Последовательность $a_n = \left(1 + \dfrac{1}{n}\right)^{n+1}$ ограничена, монотонно убывает и также сходится к числу e.

4. Последовательность $a_n = \sqrt[n]{2}$ ограничена и монотонно убывает, следовательно, сходится. Вообще последовательность $a_n = \sqrt[n]{a}$ сходится, если $a > 0$, но при $0 < a < 1$ является не убывающей, а возрастающей.

5. Последовательность $a_n = \sqrt[n]{n}$, начиная с третьего члена, монотонно убывает, ограничена, ее предел равен 1.

ДЕЙСТВИЯ НАД СХОДЯЩИМИСЯ ПОСЛЕДОВАТЕЛЬНОСТЯМИ

Следующие теоремы о действиях над сходящимся последовательностями очень часто облегчают нахождение пределов.

1. Если $\lim a_n = A$ и $\lim b_n = B$, то
$\lim (a_n + b_n) = A + B$, т. е. предел суммы (существует и) равен сумме пределов.

2. Если $\lim a_n = A$ и $\lim b_n = B$, то
$\lim (a_n \cdot b_n) = A \cdot B$,
т. е. предел произведения (существует и) равен произведению пределов.

3. Если $\lim a_n = A$ и $\lim b_n = B$, то
$\lim (a_n - b_n) = A - B$, т. е. предел разности (существует и) равен разности пределов.

4. Если $\lim a_n = A$, $\lim b_n = B$ и $B \neq 0$, то

$$\lim \left(\frac{a_n}{b_n}\right) = \frac{A}{B},$$

т. е. если предел знаменателя не равен 0, то предел отношения (существует и) равен отношению пределов.

5. Если $a_n \leqq b_n \leqq c_n$ и $\lim\limits_{n\to\infty} a_n = \lim\limits_{n\to\infty} c_n = A$, то

$$\lim_{n\to\infty} b_n = A.$$

6. Если $a_n \leqq b_n$ и $\lim\limits_{n\to\infty} a_n = A$, $\lim\limits_{n\to\infty} b_n = B$, то

$$A \leqq B.$$

Доказательства этих шести теорем опираются на правила действий с абсолютными величинами и неравенствами. Применим эти теоремы для решения нескольких примеров.

1. В примерах предыдущего параграфа упоминалась последовательность $a_n = x^n$. Мы установили тогда, что при $0 < x < 1$ последовательность сходится. Найдем теперь ее предел. Из самого понятия сходимости очевидно, что, если $a_n \to l$, то и $a_{n+1} \to l$. В нашем случае имеем:

$x^n \to l \quad x^{n+1} \to l$, в то же время

$$\begin{array}{ccc} x^{n+1} = & x & \cdot\ x^n \\ \downarrow & \downarrow & \downarrow \\ l = & x & \cdot\ l \end{array}$$

предел произведения равен произведению пределов, следовательно,

$l(1-x)=0$, т. е. $l=0$.

Отсюда следует, что, если $x>1$, то $x^n \to \infty$, т. к. для числа, большего 1, обратным является положительное число, меньшее, чем 1.

2. Как было установлено, последовательность $a_n = \sqrt[n]{2}$ сходится, но ее предел также не был найден. Обозначим предел последовательности $\{a_n\}$ через l. Очевидно, что l будет также пределом последовательности, которую мы получим, взяв из последовательности $\{a_n\}$ лишь члены с четными индексами, т. е.

$b_n = a_{2n} = \sqrt[2n]{2} \quad \lim b_n = l$, в то же время

$$\begin{array}{ccc} a_n = & b_n & \cdot\ b_n, \\ \downarrow & \downarrow & \downarrow \\ l = & l & \cdot\ l. \end{array}$$

отсюда

Поскольку l отлично от 0,

$l=1.$

3.
$$a_n = \frac{3n^2+5n+2}{4n^2+3n+8} = \frac{3+\dfrac{5}{n}+\dfrac{2}{n^2}}{4+\dfrac{3}{n}+\dfrac{8}{n^2}}, \text{ но}$$

$$\left(3+\frac{5}{n}+\frac{2}{n^2}\right) \to 3 \text{ и } \left(4+\frac{3}{n}+\frac{8}{n^2}\right) \to 4, \text{ поэтому } a_n \to \frac{3}{4}.$$

4. $a_n = \sqrt{n+1}-\sqrt{n}$. Умножая и деля на сопряженный множитель, получаем:

$$a_n = \frac{(\sqrt{n+1}-\sqrt{n})(\sqrt{n+1}+\sqrt{n})}{\sqrt{n+1}+\sqrt{n}} = \frac{1}{\sqrt{n+1}+\sqrt{n}}, \text{ т. е. } a_n \to 0.$$

ОБЩИЙ КРИТЕРИЙ СХОДИМОСТИ КОШИ

До сих пор мы ознакомились лишь с критерием сходимости для монотонных последовательностей. Общий критерий сходимости Коши распространяется на произвольные числовые последовательности:

последовательность $\{a_n\}$ сходится тогда и только тогда, когда для любого числа $\varepsilon>0$ найдется такое число N, что $|a_n - a_m| < \varepsilon$ лишь только $n>N$ и $m>N$.

Смысл критерия: сходимость последовательности означает, что ее члены отличаются друг от друга сколь угодно мало, лишь только их индексы достаточно велики. Критерий сходимости Коши эквивалентен принципу вложенных интервалов (см. стр. 385).

ПРЕДЕЛЬНЫЕ ТОЧКИ ПОСЛЕДОВАТЕЛЬНОСТЕЙ. ПОДПОСЛЕДОВАТЕЛЬНОСТИ

Рассмотрим опять последовательность

$$a_n = (-1)^n.$$

Мы уже установили, что эта последовательность не сходится. Однако, если рассмотреть лишь члены с четными индексами, то мы получим последовательность

$$b_n = a_{2n} = 1\,,$$

которая сходится.

Рассматривая члены с нечетными индексами, мы также получим сходящуюся последовательность:

$$c_n = a_{2n-1} = -1\,.$$

Таким же образом можно ввести следующее общее понятие: если $\{m\}_n$ — возрастающая последовательность натуральных чисел, то

$$\{a_{m_n}\}$$

является *подпоследовательностью* последовательности $\{a_n\}$. Если последовательность $\{a_n\}$ сходится к числу l, то к тому же самому числу сходится и всякая ее подпоследовательность. Однако, как мы только что видели, сходящиеся подпоследовательности могут быть и у не сходящейся последовательности. *Предельной точкой последовательности* $\{a_n\}$ называется всякое число, к которому сходится некоторая ее подпоследовательность. Например, предельными точками последовательности $a_n = (-1)^n$ являются, как мы уже видели, 1 и -1. Очевидно, что других предельных точек у этой последовательности нет.

Если последовательность $\{a_n\}$ ограничена сверху, то среди ее предельных точек есть наибольшая. Ее обозначают следующим образом:

$$\overline{\lim_{n\to\infty}}\, a_n \qquad \text{(limes superior — верхний предел)}\,.$$

Если последовательность $\{a_n\}$ не ограничена сверху, то существует ее подпоследовательность, стремящаяся к ∞. В этом случае говорят, что $\overline{\lim\limits_{n\to\infty}}\, a_n = \infty$.

Аналогичным образом определяется и наименьшая предельная точка последовательности:

$$\varliminf_{n\to\infty} a_n \qquad \text{(limes inferior — нижний предел)}.$$

Очевидно, что последовательность $\{a_n\}$ сходится и ее пределом является l тогда и только тогда, когда

$$\varliminf a_n = \varlimsup a_n = l\,.$$

Число $\varlimsup a_n$ обладает еще и следующим свойством: если $\varlimsup a_n < M$, то члены последовательности, индексы которых достаточно велики, меньше, чем M; если же $M = \varlimsup a_n$, то существует бесконечно много членов последовательности, больших, чем M. Аналогичным свойством обладает и $\varliminf a_n$.

ПРЕДЕЛЬНЫЕ ТОЧКИ МНОЖЕСТВ И ИХ СВЯЗЬ С ПОСЛЕДОВАТЕЛЬНОСТЯМИ

Если H — некоторое множество точек на прямой и l — предельная точка множества H, то существует последовательность, состоящая из различных точек множества H и сходящаяся к l. В самом деле, пусть a_1 — точка множества H, отличающаяся от l, но меньше, чем на 1; a_2 — точка множества H, отличающаяся от l, но меньше, чем на $\frac{1}{2}\,|\,l - a_1\,|$; a_3 — точка множества H, отличающаяся от l, но меньше, чем на $\frac{1}{2}\,|\,l - a_2\,|$. Продолжая этот процесс дальше, мы получим по порядку члены некоторой последовательности $\{a_n\}$, сходящейся к l.

Обратно, если последовательность $\{a_n\}$ состоит из различных точек множества H и сходится к l, то l — предельная точка множества H.

БЕСКОНЕЧНЫЕ РЯДЫ

Вернемся опять к бесконечным периодическим десятичным дробям, о которых шла речь при введении вещественных чисел, и рассмотрим, например, дробь

$$0{,}777\ldots,$$

а также следующую последовательность:

$$s_1 = 0{,}7, \quad s_2 = 0{,}77, \ldots, \quad s_n = 0{,}\underbrace{7}_{1}\underbrace{7}_{2}\ldots\underbrace{7}_{n}$$

Это можно записать иначе:

$$s_1 = \frac{7}{10}, \qquad s_2 = \frac{7}{10} + \frac{7}{10^2}, \ldots, \qquad s_n = \frac{7}{10} + \frac{7}{10^2} + \ldots + \frac{7}{10^n}.$$

s_n является суммой первых n членов геометрической прогрессии, первый член которой $a_1 = \frac{7}{10}$, а знаменатель $q = \frac{1}{10}$. Используя формулу для суммы n членов геометрической прогрессии, получаем:

$$s_n = \frac{\frac{7}{10}\left(1 - \frac{1}{10^n}\right)}{1 - \frac{1}{10}} = \frac{7}{9}\left(1 - \frac{1}{10^n}\right).$$

Отсюда $\lim s_n = \frac{7}{9}$.

В результате мы получили, что $0{,}777\ldots = \frac{7}{9}$, и тем самым преобразовали бесконечную периодическую десятичную дробь в обыкновенную. Перепишем это иначе:

$$0{,}777\ldots = \frac{7}{10} + \frac{7}{100} + \frac{7}{1000} + \ldots = \frac{7}{9}.$$

Записанное равенство не означает, разумеется, что мы складываем бесконечно много чисел и в результате получаем $\frac{7}{9}$. Бесконечное число суммирований нельзя произвести. Речь идет о том, что $\frac{7}{9}$ является числом, от которого сумма отличается сколь угодно мало, если сложить достаточно много членов.

Рассмотрим теперь в более общем виде последовательность $\{s_n\}$ частичных сумм геометрического ряда $\sum\limits_1^{\infty} a_n$, получаемого из геометрической прогрессии, когда $q \neq 1$.

$$s_n = a_1 \cdot \frac{1 - q^n}{1 - q} = \frac{a_1}{1 - q} - \frac{a_1}{1 - q} \cdot q^n.$$

Величина $\frac{a_1}{1 - q}$ — постоянная, не зависящая от n. Если $|q| < 1$, то $q^n \to 0$, поэтому

$$s = \lim_{n \to \infty} s_n = \frac{a_1}{1 - q}, \qquad \text{если} \qquad |q| < 1.$$

При других значениях q последовательность $\{s_n\}$ не сходится.

Используя принятую терминологию, будем говорить, что бесконечный геометрический ряд сходится, если $|q| < 1$, и его сумма

$$s = \lim_{n\to\infty} s_n = \frac{a_1}{1-q}.$$

Таким образом, под суммой бесконечного геометирического ряда мы понимаем предел последовательности его частичных сумм.

Пример геометрического ряда подводит нас к общему понятию бесконечного числового ряда. Пусть дана некоторая последовательность $\{a_n\}$. Образуем из нее последовательность $\{s_n\}$ частичных сумм:

$$s_n = a_1 + \ldots + a_n = \sum_{k=1}^{n} a_k .$$

Если последовательность $\{s_n\}$ сходится, и ее пределом является s, то будем говорить, что бесконечный ряд

$$a_1 + a_2 + \ldots + a_n + \ldots = \sum_{k=1}^{\infty} a_k$$

сходится, и его суммой является s.

Таким образом, сумма бесконечного ряда получается не в результате суммирования всех членов, а как предел последовательности частичных сумм ряда. Использованные выше обозначения:

$$\sum_{k=1}^{n} a_k \quad \text{и} \quad \sum_{k=1}^{\infty} a_k$$

читают как «сумма a_k по k от 1 до n» и, соответственно, «сумма a_k по k от 1 до ∞». Символ Σ означает суммирование по указанному индексу.

ОБЩИЙ КРИТЕРИЙ СХОДИМОСТИ ДЛЯ БЕСКОНЕЧНЫХ РЯДОВ

Применяя к последовательности частичных сумм критерий Коши, получим необходимое и достаточное условие сходимости бесконечного ряда: бесконечный ряд Σa_n сходится тогда и только тогда, когда для любого $\varepsilon > 0$ найдется такое $N(\varepsilon)$, что

$$|s_n - s_m| < \varepsilon \quad \text{лишь только} \quad n > N \quad \text{и} \quad m > N,$$

$$\text{причем} \quad |s_n - s_m| = |a_{m+1} + a_{m+2} + \ldots + a_n| < \varepsilon \text{ и } n > m.$$

Рассматривая специальный случай, когда $n = m + 1$, получаем следующее чрезвычайно простое необходимое условие сходимости бесконечного ряда: бесконечный ряд Σa_n может сходиться только если $a_n \to 0$.

Гармонический ряд. Условие $a_n \to 0$ не является достаточным для сходимости ряда. В этом можно убедиться, рассмотрев так называемый гармонический ряд:

$$1 + \frac{1}{2} + \frac{1}{3} + \ldots + \frac{1}{n} + \ldots = \sum_{n=1}^{\infty} \frac{1}{n},$$

который является расходящимся. В самом деле, применяя общий критерий сходимости Коши, при $n = 2m$ получаем:

$$s_{2m} - s_m = \frac{1}{m+1} + \frac{1}{m+2} + \ldots + \frac{1}{2m} > \frac{1}{2m} + \ldots + \frac{1}{2m} = \frac{1}{2}$$

(неравенство получается из-за того, что при замене каждого слагаемого на наименьшее из всех сумма уменьшается). Отсюда видно, что для $\varepsilon = \frac{1}{2}$ условия критерия Коши не выполняются.

РЯДЫ С ПОЛОЖИТЕЛЬНЫМИ ЧЛЕНАМИ

Если члены ряда $\sum_1^\infty a_n$ — положительные числа, то последовательность $\{s_n\}$ монотонно возрастает. Таким образом получаем следующий критерий: бесконечный ряд Σa_n с $a_n \geqq 0$ сходится тогда и только тогда, когда последовательность его частичных сумм ограничена.

На этом критерии основан признак, который можно с успехом применять для установления сходимости (расходимости) рядов, основываясь на поведении уже известных рядов. Это так называемый *признак сравнения:*

если знакоположительный ряд $\sum\limits_{n=1} c_n$ сходится, и $0 < a_n \leqq c_n$, то ряд $\sum\limits_{n=1}^{\infty} a_n$ также сходится; если знакоположительный ряд $\sum\limits_{n=1}^{\infty} d_n$ расходится, и $b_n \geqq d_n$, то расходится и ряд $\sum\limits_{n=1}^{\infty} b_n$.

Рассмотрим несколько случаев применения признака сравнения.

1. Рассмотрим следующий пример:

$$\frac{1}{1 \cdot 2} + \frac{1}{2 \cdot 3} + \frac{1}{3 \cdot 4} + \ldots = \sum_{k=1}^{\infty} \frac{1}{k(k+1)}.$$

Используя тождество

$$\frac{1}{k} - \frac{1}{k+1} = \frac{1}{k(k+1)},$$

частичные суммы ряда можно представить в виде:

$$s_n = \frac{1}{1\cdot 2} + \frac{1}{2\cdot 3} + \ldots + \frac{1}{n(n+1)} =$$
$$= 1 - \frac{1}{2} + \frac{1}{2} - \frac{1}{3} + \ldots + \frac{1}{n} - \frac{1}{n+1} = 1 - \frac{1}{n+1}.$$

Поскольку $\lim\limits_{n\to\infty}\left(1 - \frac{1}{n+1}\right) = 1$, наш ряд сходится, его сумма равна 1.

Теперь можно использовать этот сходящийся ряд для доказательства сходимости ряда

$$1 + \frac{1}{2^2} + \frac{1}{3^2} + \ldots = \sum_{n=1}^{\infty} \frac{1}{n^2}.$$

Если сравнивать два ряда почленно, то признак сравнения применить не удается. Однако если отбросить первый член и рассмотреть ряд

$$\frac{1}{2^2} + \frac{1}{3^2} + \ldots = \sum_{k=1}^{\infty} \frac{1}{(k+1)^2},$$

то признак уже может быть использован, т. к.

$$\frac{1}{k(k+1)} > \frac{1}{(k+1)^2},$$

поскольку из двух дробей больше та, знаменатель которой меньше.

Отбрасывание первого члена (и вообще конечного числа членов) ряда не влияет на его сходимость. Естественно, что сумма ряда при этом меняется, но мы и не вычисляли сумму ряда $\sum\limits_{k=1}^{\infty} \frac{1}{k^2}$, а лишь доказали его сходимость.

Рассмотрим теперь ряд

$$1 + \frac{1}{2^\alpha} + \frac{1}{3^\alpha} + \ldots = \sum_{k=1}^{\infty} \frac{1}{k^\alpha}.$$

Чем больше α, тем больше знаменатель дробей, тем, следовательно, меньше члены ряда. Выше мы получили сходимость в случае $\alpha = 2$, отсюда следует, что ряд сходится и для произвольного $\alpha > 2$, т. е. сходится, например, ряд

$$1 + \frac{1}{2^2\sqrt{2}} + \frac{1}{3^2\sqrt{3}} + \ldots = \sum_{n=1}^{\infty} \frac{1}{n^2\sqrt{n}} = \sum_{n=1}^{\infty} \frac{1}{n^{5/2}}.$$

При $\alpha = 1$ мы назвали ряд гармоническим и доказали, что он расходится. Отсюда следует, что ряд расходится и при всех $\alpha < 1$. Так, например, ряд

$$1 + \frac{1}{\sqrt{2}} + \frac{1}{\sqrt{3}} + \ldots = \sum_{n=1}^{\infty} \frac{1}{n^{1/2}}$$

расходится.

2. В предыдущем примере осталось невыясненным поведение ряда $\sum \frac{1}{n^\alpha}$, когда $1 < \alpha < 2$. В этом случае на помощь может прийти следующая простая теорема:

пусть $\sum_{n=1}^{\infty} a_n$ — ряд с положительными монотонно убывающими членами, т. е.

$$a_1 \geqq a_2 \geqq \ldots \geqq a_n \geqq a_{n+1} \geqq \ldots$$

Тогда этот ряд сходится (или расходится) одновременно со следующим:

$$a_1 + 2a_2 + 4a_4 + 8a_8 + \ldots = \sum_{n=1}^{\infty} 2^n \cdot a_{2^n}.$$

Теперь вернемся к ряду $\sum_{n=1}^{\infty} \frac{1}{n^\alpha}$. Его члены монотонно убывают, следовательно, применима предыдущая теорема. Значит, вместо исходного ряда можно рассмотреть ряд $\sum 2^n \cdot \frac{1}{2^{n\alpha}}$. Запишем его иначе: $\sum 2^{n(1-\alpha)} = \sum [2^{1-\alpha}]^n$. Мы получили геометрический ряд со знаменателем $2^{1-\alpha}$. Геометрический ряд сходится, если абсолютная величина знаменателя меньше 1. Последнее выполняется тогда и только тогда, когда $\alpha > 1$.

Подведем итог: $\sum_{n=1} \frac{1}{n^\alpha}$ сходится при $\alpha > 1$ и расходится во всех остальных случаях.

ПРИЗНАКИ ДАЛАМБЕРА И КОШИ

Из признака сравнения могут быть получены часто используемые признаки Даламбера и Коши. Оба признака получаются путем сравнения с геометрическим рядом.

Признак Даламбера: если для знакоположительного ряда $\sum a_n$ при всех достаточно больших n выполняется неравенство

$$\frac{a_{n+1}}{a_n} < q < 1,$$

то ряд сходится. Если же для всех достаточно больших n

$$\frac{a_{n+1}}{a_n} \geqq 1,$$

то ряд расходится. Этот признак можно сформулировать иначе: если существует

$$\lim_{n\to\infty} \frac{a_{n+1}}{a_n} = l,$$

и $l < 1$, то ряд сходится, если же $l > 1$, то расходится. Если последовательность $\left\{\frac{a_{n+1}}{a_n}\right\}$ не сходится, то признак формулируется следующим образом: если $\overline{\lim}\, \frac{a_{n+1}}{a_n} < 1$, то ряд сходится, если $\underline{\lim}\, \frac{a_{n+1}}{a_n} > 1$ — расходится.

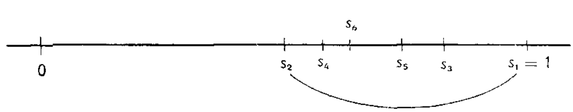
Признак Коши: если для знакоположительного ряда $\Sigma\, a_n$ при всех достаточно больших n выполняется неравенство $\sqrt[n]{a_n} < l < 1$, то ряд сходится, если же $\sqrt[n]{a_n} \geqq 1$, то расходится. Здесь также возможна формулировка «на языке пределов»: если $\overline{\lim}\, \sqrt[n]{a_n} < 1$, то ряд сходится, если же $\overline{\lim}\, \sqrt[n]{a_n} > 1$, то расходится.

ПРИЗНАК ЛЕЙБНИЦА ДЛЯ ЗНАКОПЕРЕМЕННЫХ РЯДОВ

До сих пор мы занимались рядами с положительными членами. Рассмотрим теперь ряд

$$1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \ldots = \sum (-1)^{n-1} \cdot \frac{1}{n}.$$

В том, что этот ряд сходится, можно наглядно убедиться следующим образом: будем изображать частичные суммы ряда точками на числовой оси. Начнем с того, что установим острие карандаша в точке 0. Чтобы пе-



Изображение на числово оси частичных сумм ряда $\sum (-1)^{n-1} \cdot \frac{1}{n}$

рейти к точке s_1, нужно передвинуть карандаш на единицу расстояния вправо. Поскольку a_2 отрицательно, то, чтобы отметить теперь точку, соответствующую сумме s_2, нужно передвинуть карандаш влево, но уже на меньшее расстояние, чем в первый раз. Чтобы попасть в точку s_3, карандаш из точки s_2 нужно сдвинуть опять вправо, но на меньшее расстояние, чем перед этим, и т. д. Движение карандаша при этом будет напоминать затухающее колебание, величина амплитуды которого стремится к 0, т. к. равна $\frac{1}{n}$. Точка оси, вокруг которой колебание происходит и при неограниченном уменьшении амплитуды, является пределом.

От приведенной наглядной иллюстрации очень легко перейти к строгому изложению. Последовательность частичных сумм с нечетными индексами монотонно убывает и ограничена, следовательно, сходится. Обозначим ее предел через s'. Последовательность частичных сумм с четными индексами монотонно возрастает и также ограничена, ее предел обозначим через s''. Поскольку

$$s_{2n+1} - s_{2n} = a_{2n+1}$$

$$\lim s_{2n+1} - \lim s_{2n} = s' - s'' = \lim a_{2n+1} = 0\,,$$

получаем, что

$$s' = s'',$$

т. е. все частичные суммы ряда стремятся к одному и тому же пределу.

При изложении доказательства мы нигде не использовали конкретных значений членов ряда. Мы опирались лишь на то, что ряд *знакочередующийся* (колебание), что члены ряда по абсолютной величине монотонно убывают (затухание) и что $\lim a_n = 0$ (существует точка, вокруг которой колебание происходит и при неограниченном уменьшении амплитуды). Теперь можно сформулировать *признак Лейбница*:

если абсолютные величины членов знакочередующегося ряда монотонно убывая стремятся к 0, то ряд сходится.

РЯДЫ С ЛЮБЫМИ ВЕЩЕСТВЕННЫМИ ЧЛЕНАМИ

Рассмотрим два сходящиеся ряда:

$$\sum (-1)^{n-1}\cdot\frac{1}{n^2} = 1 - \frac{1}{4} + \frac{1}{9} - \frac{1}{16} + \ldots$$

$$\sum (-1)^{n-1}\cdot\frac{1}{n} = 1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \ldots$$

Первый ряд останется сходящимся и после замены каждого члена на его абсолютную величину, т. к. получаемый при этом ряд $\sum \frac{1}{n^2}$, как мы видели, сходится. Однако, произведя аналогичную замену во втором ряду, мы получим гармонический ряд, который, как известно, расходится. Оба эти примера подводят нас к очень важному понятию абсолютной сходимости.

Ряд Σa_n называется *абсолютно сходящимся*, если сходится ряд $\Sigma |a_n|$.

Нетрудно убедиться в том, что всякий абсолютно сходящийся ряд сходится, однако, как показывает второй пример, не всякий сходящийся ряд сходится абсолютно. Тот факт, что абсолютная сходимость является достаточным условием сходимости, дает возможность применять к рядам с любыми вещественными членами в качестве достаточных условий те признаки сходимости, которые были установлены для рядов с положительными членами. Если, например,

$$\left|\frac{a_{n+1}}{a_n}\right| < q < 1 \qquad \text{или} \qquad \overline{\lim}\sqrt[n]{|a_n|} < 1,$$

то ряд Σa_n абсолютно сходится и, следовательно, сходится.

ДЕЙСТВИЯ С БЕСКОНЕЧНЫМИ РЯДАМИ

Суммирование бесконечных рядов. Если бесконечные ряды Σa_n и Σb_n сходятся, и их суммы равны, соответственно, A и B, то ряд $\Sigma(a_n + b_n)$ также сходится, и его сумма равна $A \pm B$. Это следует из того, что предел суммы равен сумме пределов. Аналогично доказывается, что $\Sigma ca_n = cA$, где c — некоторая постоянная.

Таким образом, сходящиеся бесконечные ряды можно почленно складывать и вычитать. Однако при вычислении произведения уже в случае конечных сумм недостаточно перемножить первый член с первым, второй — со вторым и т. д. Необходимо каждый член первой суммы перемножить с каждым членом второй. Для этого рассмотрим так называемое *произведение Коши* двух рядов:

$$(a_0 + a_1 + a_2 \ldots)(b_0 + b_1 + b_2 + \ldots) = a_0b_0 + (a_0b_1 + a_1b_0) + (a_0b_2 + \\ + a_1b_1 + a_2b_0) + (a_0b_3 + a_1b_2 + a_2b_1 + a_3b_0) + \ldots$$

Закон, по которому образуются члены этого ряда, гарантирует, что каждый член первого ряда будет перемножен с каждым членом второго. В самом деле, члены ряда сгруппированы таким образом, что в первую группу попадают произведения тех членов перемножаемых рядов, сумма индексов которых равна 0, во второй группе сумма индексов равна 1, в третьей — 2 и т. д. Таким образом записать произведение рядов, зная этот закон, нетрудно. Однако выяснить, сходится ли получаемый ряд, не так просто, как в случае суммирования рядов. Если оба исходных ряда сходятся абсолютно, то можно показать, что и их произведение является сходящимся рядом, сумма которого равна произведению сумм «сомножителей».

Говоря о сумме бесконечного ряда, естественно поставить вопрос, выполняются ли в этом случае ассоциативный и коммутативный законы сложения. Иначе говоря, можно ли произвольным образом группировать члены ряда, расставлять или опускать скобки, а также можно ли произвольно переставлять члены ряда.

Группировать члены ряда, расставляя скобки, очевидно, можно — это не влияет ни на сходимость, ни на сумму ряда, т. к. все частичные суммы получаемого ряда являются частичными суммами и для исходного ряда. Опускать скобки следует уже осмотрительно. В самом деле,

$$0 + 0 + 0 + \ldots = (1 - 1) + (1 - 1) + (1 - 1) + \ldots = 0.$$

Однако, опустив скобки, мы получим следующий ряд:

$$1 - 1 + 1 - 1 + \ldots$$

Этот ряд уже расходится, т. к. $s_1 = s_3 = s_5 = \ldots, 1$, в то же время

$$s_2 = s_4 = s_6 = \ldots = 0.$$

Значит, скобки нельзя опускать произвольным образом. Однако если ряд, получаемый после опускания скобок, сходится, то скобки можно опустить.

Проверка выполнения коммутативного закона сложения приводит к неожиданным результатам. Мы уже видели, что ряд

$$1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \ldots \quad \text{сходится.}$$

Обозначим его сумму через S. Сгруппируем члены ряда следующим образом:

$$\left(1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4}\right) + \left(\frac{1}{5} - \frac{1}{6} + \frac{1}{7} - \frac{1}{8}\right) + \\ + \left(\frac{1}{9} - \frac{1}{10} + \frac{1}{11} - \frac{1}{12}\right) + \ldots$$

Согласно этой группировке,

$$S = \sum_{n=0}^{\infty} \left(\frac{1}{4n+1} - \frac{1}{4n+2} + \frac{1}{4n+3} - \frac{1}{4n+4}\right).$$

Если же сгруппировать члены ряда попарно:

$$\left(1 - \frac{1}{2}\right) + \left(\frac{1}{3} - \frac{1}{4}\right) + \ldots, \quad \text{то}$$

$$S = \sum_{n=0}^{\infty} \left(\frac{1}{2n+1} - \frac{1}{2n+2}\right).$$

Умножив все члены последнего ряда на $\frac{1}{2}$, получаем:

$$\frac{1}{2}S=\sum_{n=0}^{\infty}\left(\frac{1}{4n+2}-\frac{1}{4n+4}\right).$$

Полученный ряд сложим почленно с рядом, где члены группировались по четыре:

$$\sum_{n=0}^{\infty}\left(\frac{1}{4n+1}-\frac{1}{4n+2}+\frac{1}{4n+3}-\frac{1}{4n+4}+\frac{1}{4n+2}-\frac{1}{4n+4}\right).$$

Упрощая выражение в скобках, получаем:

$$\frac{3}{2}S=\sum_{n=0}^{\infty}\left(\frac{1}{4n+1}+\frac{1}{4n+3}-\frac{2}{4n+4}\right)=$$
$$=\sum_{n=0}^{\infty}\left(\frac{1}{4n+1}+\frac{1}{4n+3}-\frac{1}{2n+2}\right).$$

Распишем полученный ряд:

$$\frac{3}{2}S=1+\frac{1}{3}-\frac{1}{2}+\frac{1}{5}+\frac{1}{7}-\frac{1}{4}+\ldots,$$

в то же время

$$S=1-\frac{1}{2}+\frac{1}{3}-\frac{1}{4}+\ldots$$

Неожиданным является то, что ряд с суммой $\frac{3}{2}S$ состоит из тех же самых членов, что и ряд с суммой S, только записанных в другом порядке. (Нетрудно показать, что, расписывая ряд $\frac{3}{2}S$, мы имели право опустить скобки, однако, здесь мы не будем на этом останавливаться).

Переставляя члены того же самого ряда подходящим образом, можно получить и не две разные суммы. Задав наперед произвольное число, можно переставить члены этого ряда таким образом, что сумма будет равна как раз заданному числу. Можно совершить перестановку и так, что частичные суммы ряда будут стремиться к ∞, можно и так, что к $-\infty$. Например, ряд $1+\frac{1}{3}+\frac{1}{5}+\ldots$ расходится, т. к. в противном случае сходился бы и ряд $\frac{1}{2}+\frac{1}{4}+\frac{1}{6}+\ldots$, каждый член которого меньше соответствующего члена первого ряда, но тогда сходился бы и гармонический ряд, получаемый при суммировании обоих рядов. Ряд

$$\frac{1}{2}+\frac{1}{4}+\frac{1}{6}+\ldots=\frac{1}{2}\left(1+\frac{1}{2}+\frac{1}{3}+\ldots\right)$$

также расходится.

Вообще, если бесконечный ряд состоит из положительных и отрицательных членов, которые в отдельности образуют расходящиеся ряды, то исходный ряд может расходиться и может сходиться, причем сумма его может оказаться равной произвольному числу в зависимости от того, в каком порядке записаны члены ряда.

Безусловная и условная сходимость. Если положительные члены ряда образуют расходящийся, а отрицательные члены — сходящийся ряд, то частичные суммы исходного ряда стремятся к ∞. Если же положительные члены образуют сходящийся ряд, а отрицательные — расходящийся, то частичные суммы исходного ряда стремятся к $-\infty$. Осталось разобрать случай, когда и положительные, и отрицательные члены образуют сходящиеся ряды. В этом случае сходимость и сумма исходного ряда не зависят от порядка, в котором записаны его члены. Такой ряд называют *безусловно сходящимся*, в отличие от рядов первого типа, сходимость которых зависит от порядка записи членов и которые называют поэтому *условно сходящимися.*

Понятие безусловной сходимости раскрывается в следующей теореме: *бесконечный ряд Σa_n сходится безусловно тогда и только тогда, когда абсолютно сходится.*

СТЕПЕННЫЕ РЯДЫ

Мы уже познакомились (правда, не в такой форме) с соотношением

$$1+x+x^2+\ldots=\sum_{n=0}^{\infty}x^n=\frac{1}{1-x}.$$

Оно означает, что геометрический ряд Σx^n сходится, если $|x|<1$, и его сумма равна $\dfrac{1}{1-x}$. При всех остальных значениях x ряд расходится. Будем говорить, что областью сходимости ряда функций (все члены ряда — функции переменного x)

$$1+x+x^2+\ldots=\sum_{n=0}^{\infty}x^n$$

является открытый интервал $(-1;\ 1)$. Рассматривая последовательность частичных сумм этого ряда, будем говорить, что областью сходимости последовательности функций

$$\{s_n(x)\}=\left\{\frac{1-x^n}{1-x}\right\}$$

является открытый интервал $(-1;\ 1)$.

Введем общее понятие. Если $\{f_n(x)\}$ — последовательность функций, областью определения которых является некоторое множество H, то для произвольного элемента a множества H после подстановки мы получим числовую последовательность $\{f_n(a)\}$. Если эта числовая последовательность сходится, то будем говорить, что a принадлежит области сходимости последовательности функций $\{f_n(x)\}$. Итак, областью сходимости последовательности функций называется множество, состоящее из всех тех точек, в которых последовательность сходится. В точках области сходимости определена предельная функция

$$f(x) = \lim_{n\to\infty} f_n(x) .$$

В таком же самом смысле мы говорим об области сходимости ряда функций

$$s_0(x) + s_1(x) + \ldots = \sum_{n=0}^{\infty} s_n(x)$$

и о его сумме

$$S(x) = \sum_{n=0}^{\infty} s_n(x) .$$

Ряд вида $c_0 + c_1x + c_2x^2 + \ldots = \sum\limits_{n=0}^{\infty} c_n x^n$ называется *степенным рядом* (с центром в нуле). Название объясняется тем, что члены ряда являются степенями x. Степенные ряды играют большую роль в математике, т. к. могут применяться при решении самых различных задач. Областью сходимости степенного ряда всегда является интервал вида $(-r\,;\ r)$ с центром в точке 0, т. е. множество точек, удаленных от начальной точки меньше, чем на r. Этот интервал называют *интервалом сходимости* степенного ряда. Можно доказать, что в точках этого интервала степенной ряд абсолютно сходится. Число, равное половине длины интервала сходимости, называют *радиусом сходимости* степенного ряда (причину такой терминологии можно понять лишь после ознакомления с теорией функций комплексного переменного). Используя признак Коши, можно получить следующую формулу для вычисления r:

$$r = \frac{1}{\overline{\lim}\sqrt[n]{|a_n|}} .$$

Если $\overline{\lim}\sqrt[n]{|a_n|} = \infty$, то радиус сходимости равен нулю, т. е. ряд сходится лишь при значении $x = 0$. Если $\overline{\lim}\sqrt[n]{|a_n|} = 0$, то считается, что $r = \infty$, т. е. ряд сходится на всей числовой прямой. Во всех точках, удаленных от точки 0 больше, чем на r, степенной ряд расходится. О поведении ряда в точках, удаленных от нуля на расстояние r, т. е. о двух крайних точках интервала сходимости нельзя сказать ничего определенного. Это видно из следующих трех примеров:

$$\sum_{n=0}^{\infty} x^n; \qquad \sum_{n=0}^{\infty} \frac{x^n}{n}; \qquad \sum_{n=0}^{\infty} \frac{x^n}{n^2}.$$

Радиусы сходимости всех трех рядов равны 1. Первый ряд расходится в точках -1 и $+1$, второй сходится в точке -1 и расходится в точке $+1$, третий ряд сходится в обеих точках.

Степенным рядом (с центром в точке a) мы называем ряд вида

$$c_0 + c_1(x-a) + c_2(x-a)^2 + \ldots = \sum_{n=0}^{\infty} c_n(x-a)^n.$$

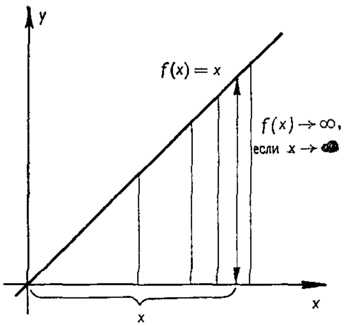
Областью сходимости этого ряда является интервал длины $2r$ с центром в точке a.

ПРЕДЕЛ ФУНКЦИИ

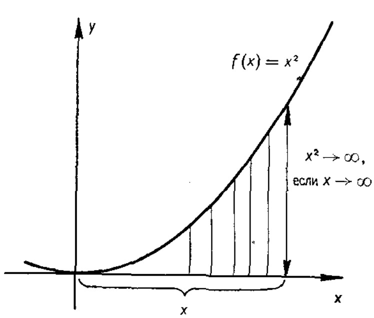
Числовая последовательность — это функция, областью определения которой является множество натуральных чисел. От понятия предела последовательности нетрудно перейти к понятию предела функции. Например, от последовательности $a_n = n$ можно перейти к функции $f(x) = x$. Утверждение, что $\lim\limits_{x\to\infty} f(x) = \infty$, становится очевидным и понятным после осмысления предела последовательности. Оно означает, что функция $f(x) = x$ принимает сколь угодно большие значения, если x достаточно велико.

Совершенно аналогично положение в случае функции $f(x) = x^2$:

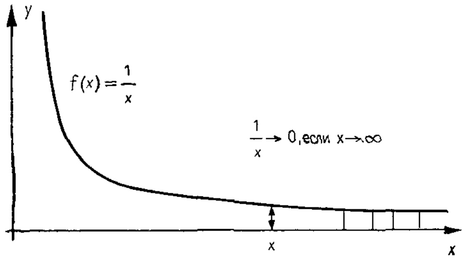
$$\lim_{x\to\infty} x^2 = \infty.$$



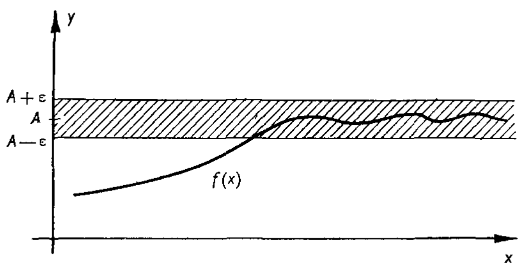
Предел функции $f(x) = x$, если $x \to \infty$, то $\lim f(x) = \infty$



Предел функции $f(x) = x^2$, если $x \to \infty$, то $\lim f(x) = \infty$



Предел функции $f(x) = \dfrac{1}{x}$, если $x \to \infty$, то $\lim f(x) = 0$



$\lim\limits_{x\to\infty} f(x) = A$, если при $x > N(\varepsilon)$ $|f(x) - A| < \varepsilon$

Это означает, что если задать произвольное число M, то

$$f(x) > M,$$

если

$$x > N(M),$$

т. е. если x больше некоторого числа N, зависящего от M. Пусть, например, $M = 10^6$, тогда $N = 10^3$. Возьмем другой пример: пусть $f(x) = \dfrac{1}{x}$. Очевидно, каков наглядный смысл того, что

$$\lim_{x\to\infty} f(x) = 0.$$

После всего сказанного совсем нетрудно сформулировать понятие предела для $f(x)$ в общем виде. Так,

$$\lim_{x\to\infty} f(x) = A$$

означает, что при достаточно больших x значения функции будут попадать в полосу, ограниченную прямыми $y = A - \varepsilon$ и $y = A + \varepsilon$, т. е.

$$|f(x) - A| < \varepsilon, \quad \text{если} \quad x > N(\varepsilon).$$

Предел функции $f(x)$ при $x \to a$ записывается как

$$\lim_{x\to a} f(x) = A$$

и означает, что $f(x)$ сколь угодно мало отличается от A, когда x находится в достаточной близости от a. Иначе говоря, если произвольно задать положительное число ε, то график кривой $y = f(x)$ окажется между прямыми $y = A - \varepsilon$ и $y = A + \varepsilon$ для всех тех значений x, которые находятся в достаточной близости от a и отличаются от самого a.

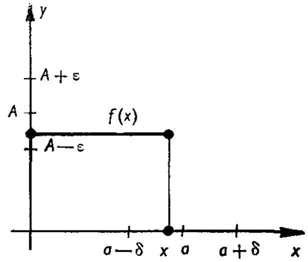
Множество, которое получится, если из ε-окрестности точки a исключить саму точку a, мы называем редуцированной ε-окрестностью точки a.

Поэтому $\lim\limits_{x\to a} f(x) = A$ означает, что для любой ε-окрестности $K_{A,\varepsilon}$ точки A найдется такая редуцированная δ-окрестность $K'_{a,\delta}$ точки a, что если x содержится в $K'_{a,\delta}$, то $f(x)$ попадает в $K_{A,\varepsilon}$. С помощью неравенств это можно выразить следующим образом:

$$|f(x) - A| < \varepsilon, \quad \text{если} \quad 0 < |x - a| < \delta(\varepsilon).$$

Предел слева и справа. Разумеется, не у всех функций существуют пределы во всех точках. Рассмотрим, например, функцию

$$y = f(x) = \begin{cases} x, & \text{если} \quad x < 2 \\ 2{,}5, & \text{если} \quad x = 2 \\ x + 1, & \text{если} \quad x > 2\,. \end{cases}$$



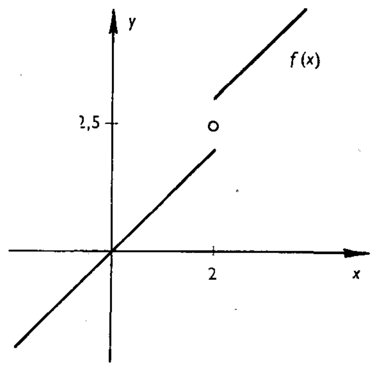
$\lim\limits_{x \to a} f(x) = A$, если при

$0 < |x - a| < \delta(\varepsilon)$

$|f(x) - A| < \varepsilon$

У этой функции не существует предела в точке $x = 2$, т. к. в какой бы близости от точки 2 мы ни оказались, для любого заданного числа найдется такое значение функции, которое отличается от него больше, чем на $\frac{1}{2}$. Однако у $f(x)$ существует предел, если x стремится к 2 слева или справа. Обозначается это следующим образом:

$$\lim_{x \to 2-0} f(x) = 2\,, \qquad \lim_{x \to 2+0} f(x) = 3\,.$$



Эта функция не имеет предела в точке $x = 2$, но имеет односторонние пределы при стремлении x к 2 слева или справа

Вообще, предел некоторой функции $f(x)$ в точке a справа:

$$\lim_{x \to a+0} f(x) = A\,,$$

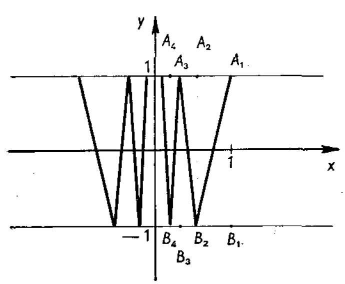
означает, что $f(x)$ сколь угодно мало отличается от A, если x находится в достаточной близости от a и больше a, т. е. для любого $\varepsilon > 0$ найдется такое $\delta > 0$, что если

$$0 < x - a < \delta\,,$$

тогда

$$|f(x) - A| < \varepsilon\,.$$

Аналогичным образом определяется предел слева: $\lim\limits_{x \to a-0} f(x)$.



Эта функция не стремится ни к какому пределу

Другой пример: возьмем систему координат и проведем прямые $y = 1$ и $y = -1$. Отметим на этих прямых последовательности точек $A_1, A_2, \ldots$ и B_1, $B_2, \ldots$ Координаты членов последовательностей пусть будут $(1\,;\,1)$, $\left(\frac{1}{2}\,;\,1\right)$, $\left(\frac{1}{3}\,;\,1\right), \ldots$ и, соответственно, $(+1\,;\,-1)$, $\left(\frac{1}{2}\,;\,-1\right)$, $\left(\frac{1}{3}\,;\,-1\right), \ldots$

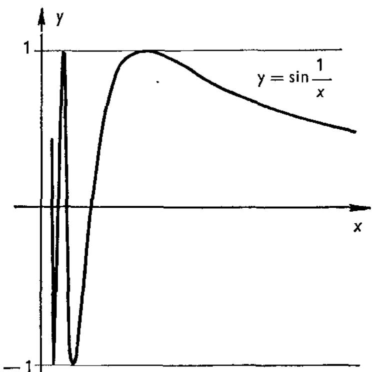
Таким образом, точка B_4, например, имеет координаты $\left(\frac{1}{4};\ -1\right)$, точка A_{10} — координаты $\left(\frac{1}{10};\ 1\right)$. После этого соединим отрезками прямых точку A_1 с точкой B_2, точку B_2 — с A_3, точку A_3 — с B_4 и т. д. Полученная ломаная линия графически задает некоторую функцию для всех тех x, для которых $0 < x < 1$.

В тех точках x, для которых $-1 < x < 0$, функцию можно определить, зеркально отразив ломаную линию относительно оси y. В точке $x = 0$ присвоим функции значение 0. Нетрудно убедиться в том, что определенная таким образом функция не стремится ни к какому пределу, когда x стремится к 0, т. к. в любой окрестности нуля всегда можно указать два значения функции, разность между которыми больше 1.

Аналогично ведет себя функция, задаваемая формулой

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

График этой функции отличается от предыдущего тем, что у него ни в одной точке нет излома.



Функция $y = \sin\frac{1}{x}$ не стремится ни к какому пределу

Бесконечные пределы функций. Из наглядных соображений нетрудно уяснить смысл следующего утверждения: функция $f(x) = \frac{1}{(x-1)^2}$ неограниченно возрастает, когда $x \to 1$. В этом случае говорят также, что функция $f(x) = \frac{1}{(x-1)^2}$ стремится к бесконечности в точке $x=1$, или, что предел функции равен ∞. Обозначают этот факт следующим образом:

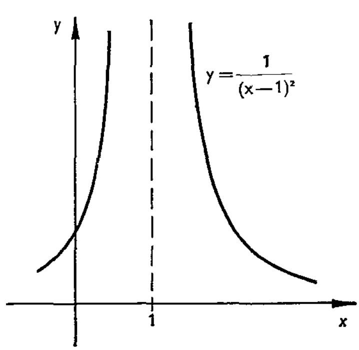
$$\lim_{x\to a} f(x) = \infty,$$

это означает, что если произвольно задать число M, то значения функции становятся больше заданного числа, лишь только x находится в достаточной близости от a. Формально:

$$f(x) > M,$$

если

$$0 < |x - a| < \delta(M).$$



При $x \to 1$ функция $f(x) = \frac{1}{(x-1)^2}$ стремится к бесконечности

Важнейшие теоремы о пределах функций (подразумевается, что все пределы существуют и конечны):

I. Предел суммы равен сумме пределов.
II. Предел разности равен разности пределов.
III. Предел произведения равен произведению пределов.
IV. Предел отношения равен отношению пределов.

Задачи, связанные с пределами функций, всегда можно свести к аналогичным задачам для последовательностей. Точнее, для того, чтобы

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

необходимо и достаточно, чтобы для любой сходящейся к a последовательности $\{x_n\}$, члены которой отличны от a, соответствующая последовательность значений функции $\{f(x_n)\}$ сходилась к A, т. е. чтобы

$$\lim_{n \to \infty} x_n = a \quad \text{влекло за собой} \quad \lim_{n \to \infty} f(x_n) = A.$$

Несколько важных пределов:

$$\lim_{x \to 0} \frac{\sin x}{x} = 1;$$

$$\lim_{n \to \infty} \left(1 + \frac{x}{n}\right)^n = e^x;$$

$$\lim_{h \to 0} \frac{\ln(x+h) - \ln x}{h} = \frac{1}{x};$$

$$\lim_{x \to \infty} \left(1 + \frac{a}{x}\right)^x = e^a.$$

Нахождение пределов.

Примеры:

1. Пусть $$y = f(x) = \frac{3x^2 + 2x + 5}{2x^2 - x + 3}.$$

Чему равен $\lim\limits_{x \to \infty} f(x)$?

Для нахождения этого предела разделим числитель и знаменатель дроби на x^2 — наибольшую степень x, имеющуюся в знаменателе.

Предел отношения равен отношению пределов. Поскольку предел числителя стал равен 3, а предел знаменателя -2, то

$$\lim_{x\to\infty} f(x) = \frac{3}{2}.$$

2.

$$\lim_{x\to 1} \frac{2x^2+3x-5}{x^2-1}.$$

Предел числителя равен 0.
Предел знаменателя равен 0.

Таким образом теорема о пределе отношения не применима. Поступим иначе: преобразуем и числитель, и знаменатель дроби в произведение. Знаменатель представляет собой разность квадратов:

$$x^2-1=(x+1)(x-1).$$

Чтобы преобразовать в произведение числитель дроби, решим уравнение

$$2x^2+3x-5=0.$$

Решениями являются

$$x_1=1 \quad \text{и} \quad x_2=-\frac{5}{2}.$$

Таким образом квадратный трехчлен $2x^2+3x-5$ преобразуется следующим образом:

$$2x^2+3x-5=2\left(x+\frac{5}{2}\right)(x-1).$$

Подставляя полученные произведении в исходную функцию, получаем:

$$\lim_{x\to 1} \frac{2x^2+3x-5}{x^2-1} = \lim_{x\to 1} \frac{2\left(x+\frac{5}{2}\right)(x-1)}{(x+1)(x-1)} = \lim_{x\to 1} \frac{2\left(x+\frac{5}{2}\right)}{x+1} = \frac{7}{2}.$$

3.

$$\lim_{x\to 0} \frac{1-\cos x}{x^2} = \lim_{x\to 0} \frac{1-\cos x}{x^2}\cdot\frac{1+\cos x}{1+\cos x} =$$

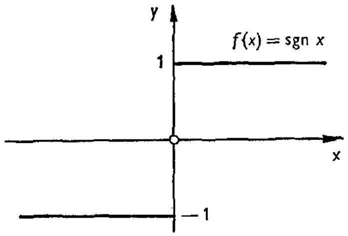
$$= \lim_{x\to 0} \frac{1-\cos^2 x}{x^2}\cdot\frac{1}{1+\cos x} = \lim_{x\to 0} \frac{\sin^2 x}{x^2}\cdot\frac{1}{1+\cos x} = \frac{1}{2}.$$

НЕПРЕРЫВНОСТЬ

Строгое определение понятия непрерывности можно дать лишь используя понятие предела.

Рассмотрим сначала несколько примеров.

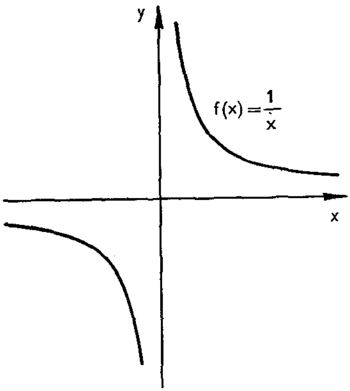
1. $y = f(x) = \operatorname{sgn} x = \begin{cases} 1, & > 0 \\ 0, \text{ если } & x = 0 \\ -1, & < 0 \end{cases}$



Функция $f(x) = \operatorname{sgn} x$ непрерывна всюду за исключением точки 0

Функция непрерывна всюду, кроме точки 0. В точке 0 у функции не существует предела, т. к. предел справа равен 1, а предел слева равен −1.

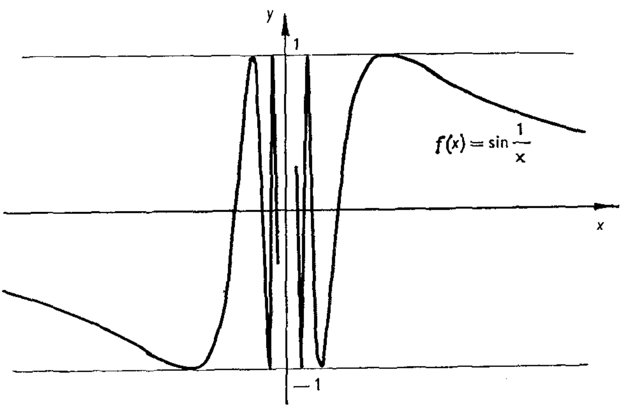
2. $y = f(x) = \begin{cases} \frac{1}{x}, \text{ если } & x \neq 0 \\ 0 & x = 0. \end{cases}$



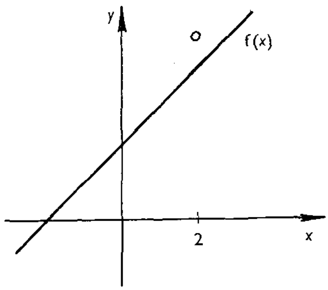
Функция $f(x) = \frac{1}{x}$ не является непрерывной в точке 0

Функция не является непрерывной в точке 0, т. к. не имеет в этой точке конечного предела ни справа, ни слева. Если x стремиться к 0 справа, то предел равен ∞, если же слева, то $-\infty$.

3. $y = f(x) = \begin{cases} \sin \frac{1}{x}, \text{ если } & x \neq 0 \\ 0 & x = 0. \end{cases}$



Функция $f(x) = \sin \frac{1}{x}$ в точке 0 не имеет предела ни справа, ни слева



Предел функции $y=f(x)$ в точке $x=2$ отличается от значения функции в этой точке

В точке 0, как мы уже видели, у функции не существует предела ни справа, ни слева.

$$4.\ y=f(x)=\begin{cases}\dfrac{x^2-4}{x-2}, & \text{если} \quad x\neq 2\\ 5 & \qquad\qquad x=2.\end{cases}$$

Здесь неприятность возникает при $x=2$. Хотя в этой точке и существует предел $\lim\limits_{x\to 2} f(x)=4$, но он не совпадает со значением функции в той же точке.

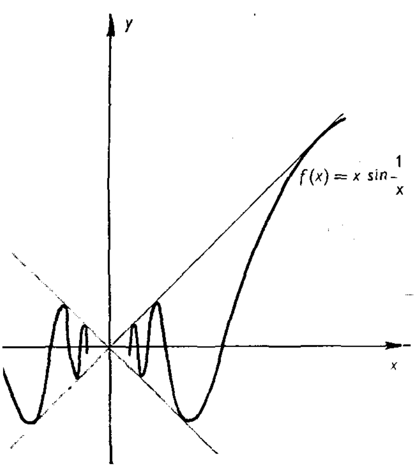
Теперь можно сформулировать определение непрерывности:

функция $f(x)$ называется непрерывной в точке a своей области определения, если $\lim\limits_{x\to a} f(x)=f(a)$, т. е. если в точке a существует конечный предел функции, и этот предел совпадает со значением самой функции в той же точке.

Это означает, что $f(x)$ сколь угодно мало отличается от $f(a)$, если x находится в достаточной близости от a.

Наглядный смысл этого понятия точно передается той элементарной формулировкой, которая была дана выше: график непрерывной функции можно нарисовать, не отрывая карандаша от бумаги. Однако понятие непрерывности, определенное сейчас, значительно шире, т. к., например, функция

$$y=f(x)=\begin{cases}x\sin\dfrac{1}{x}, & \text{если} \quad x\neq 0\\ 0 & \qquad\qquad x=0\end{cases}$$



Функция $y=x\sin\dfrac{1}{x}$ непрерывна на всей числовой оси, но ее график нельзя нарисовать

непрерывна на всей оси, но тем не менее ее график нельзя нарисовать, т. к. в промежутке между точками -1 и $+1$ функция совершает бесконечное число колебаний.

Точки разрывов функций. Если функция не является непрерывной в некоторой точке, то говорят, что в этой точке она имеет разрыв, а саму точку называют при этом *точкой разрыва* функции.

Из определения непрерывности очевидно, что точкой разрыва может оказаться или такая точка, в которой у функции не существует конечного предела, или же такая точка, в которой предел существует, но отличается от значения самой функции. Во втором случае разрыв легко устранить, если значение функции в этой точке изменить, сделав его равным величине предела. Примером может служить упомянутая выше функция

$$y = f(x) = \begin{cases} \dfrac{x^2-4}{x-2}, & \text{если } x \neq 2, \\ 5, & \phantom{\text{если }} x = 2. \end{cases}$$

Непрерывность этой функции в точке 2 можно восстановить, если изменить ее задание, положив $f(2) = 4$. Иногда подобный ход рассуждений приводит к тому, что область определения функции расширяют. Например, областью определения функции, задаваемой формулой

$$f(x) = \frac{\sin x}{x},$$

является множество всех вещественных чисел, отличных от 0. Поскольку мы знаем, что

$$\lim_{x\to 0} \frac{\sin x}{x} = 1,$$

то к области определения функции можно добавить точку $x = 0$, положив $f(0) = 1$. *Таким образом мы получим функцию, непрерывную на всей* числовой оси.

Из правил действий над пределами вытекают следующие утверждения:

1. Сумма, разность, произведение и отношение (в точке, где знаменатель не обращается в 0) непрерывных функций являются непрерывными функциями.

2. Сложная функция, составленная из двух непрерывный функций, также является непрерывной.

3. Функция, обратная монотонной непрерывной функции, сама является непрерывной.

Если принять во внимание тот факт, что функции

$$y = x, \qquad y = C, \qquad y = e^x, \qquad y = \sin x$$

непрерывны, то из перечисленных трех утверждений следует непрерывность элементарных функций.

РАВНОМЕРНАЯ НЕПРЕРЫВНОСТЬ

Непрерывность функции $f(x)$ в точке a означает, что для любого числа $\varepsilon > 0$ найдется такое число $\delta > 0$, что $|f(x) - f(a)| < \varepsilon$ лишь только $|x - a| < \delta$.

Рассмотрим, например, функцию $f(x) = x^2$ в интервале $[0; 1]$. В каждой точке этого интервала функция непрерывна. Зададим некоторое $\varepsilon > 0$, тогда

$$|x^2 - a^2| = |x - a| \cdot |x + a| \leq |x - a|(1 + a) < 2|x - a|,$$

т. е. $|f(x) - f(a)| < \varepsilon$, если $|x - a| < \dfrac{\varepsilon}{2} = \delta$.

Для заданного ε мы получаем такое δ, которое годится одновременно во всех точках области определения. Если продолжить рассмотрение функции $y = x^2$, взяв за область ее определения теперь уже всю числовую ось, то для заданного ε в каждой отдельной точке a числовой оси мы найдем такое δ, что $|f(x) - f(a)| < \varepsilon$ лишь только $|x - a| < \delta$. Однако нам не удастся найти такого δ, которое годилось бы для заданного ε одновременно во всех точках a числовой оси.

В аналогичном положении мы окажемся, если рассмотрим функцию $f(x) = \dfrac{1}{x}$ в интервале $(0, 1)$. Здесь также не удается для всякого положительного числа ε указать такое δ, которое годилось бы в каждой точке интервала $(0, 1)$. Это означает, что величина δ зависит от a, и у множества чисел δ, соответствующих при данном ε всем точкам области определения $H = (0, 1)$, не существует положительной нижней границы (если бы существовала, то эта граница была бы общим δ).

Если функция $f(x)$ непрерывна на множестве H, и для каждого $\varepsilon > 0$ найдется такое $\delta > 0$, что $|f(x) - f(a)| < \varepsilon$, лишь только $|x - a| < \delta$ для любой точки a множества H, то говорят, что функция $f(x)$ *равномерно непрерывна* на множестве H. Данному определению эквивалентна следующая формулировка:

если для любого числа $\varepsilon > 0$ найдется такое число $\delta > 0$, что для любых двух элементов x_1 и x_2 множества H неравенство $|x_1 - x_2| < \delta$ влечет за собой $|f(x_1) - f(x_2)| < \varepsilon$, то функция $f(x)$ равномерно непрерывна на множестве H.

Очень важна следующая теорема: *если функция непрерывна на ограниченном замкнутом множестве, то она и равномерно непрерывна на этом множестве* (например, функция, непрерывная в замкнутом интервале, равномерно непрерывна в этом интервале).

НЕКОТОРЫЕ ВАЖНЫЕ СВОЙСТВА НЕПРЕРЫВНЫХ ФУНКЦИЙ

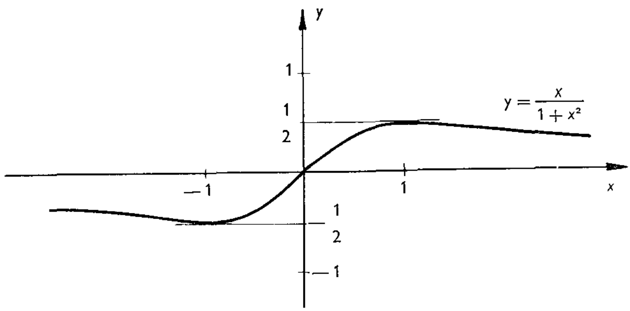
Функция $f(x)$ отображает область определения H в область значений K. Естественно поставить вопрос, что если множество H обладает какими-либо свойствами, то обладает ли теми же свойствами и множество K, являющееся его образом. Этот вопрос можно поставить и несколько иначе: какие свойства области определения сохраняются при непрерывном отображении или какие свойства множеств являются инвариантными при непрерывном отображении?

Рассмотрим несколько примеров:

Функция $y = \operatorname{tg} x$ отображает интервал $\left[-\frac{\pi}{2}; \frac{\pi}{2}\right]$ на всю числовую ось, отсюда видно, что *образ ограниченного множества при непрерывном отображении не обязательно ограничен.*

Функция $y = \operatorname{arctg} x$ отображает всю числовую ось, являющуюся замкнутым множеством, в интервал $\left(-\frac{\pi}{2}; \frac{\pi}{2}\right)$, который не замкнут. Значит, *свойство множества быть замкнутым не является инвариантным при непрерывном отображении.*

Функция $y = \frac{x}{1+x^2}$ отображает всю числовую ось в интервал $\left[-\frac{1}{2}; \frac{1}{2}\right]$, т. е. *образ открытого множества при непрерывном отображении не обязательно открыт.*



Образ открытого множества при непрерывном отображении не обязательно открыт

Исключительную важность имеет следующая теорема:

образом ограниченного замкнутого множества при непрерывном отображении является также ограниченное замкнутое множество. Следовательно, «ограниченность и замкнутость» является инвариантным свойством множеств при непрерывных отображениях. Это свойство встречается и во многих других теоремах. Так, например, мы видели, что функция, непрерывная на ограниченном замкнутом множестве, равномерно непрерывна на этом множестве. Это явилось причиной того, что для «ограниченного замкнутого» множества математики ввели специальное название: *компактное множество.*

Из теоремы следует, что

функция, непрерывная на компактном множестве, достигает своего максимума и минимума, т. е. в области определения H найдутся такие точки a_1 и a_2, что $f(a_1)$ является наибольшим, а $f(a_2)$ — наименьшим элементом области значений функции.

Действительно, ограниченное замкнутое множество содержит свои точные верхнюю и нижнюю границы, т. е. среди его элементов всегда существуют наибольший и наименьший. То, что функция, непрерывная на произвольном множестве, вообще говоря, не достигает своего максимального значения ни в одной точке, видно из следующих двух примеров: функция $y = \operatorname{tg} x$, рассматриваемая на открытом интервале $\left(-\frac{\pi}{2}; \frac{\pi}{2}\right)$, и функция $y = \operatorname{arctg} x$, рассматриваемая на всей числовой оси.

Другая важная теорема:

образ связного множества при непрерывном отображении есть связное множество.

У этой теоремы есть интересное следствие: функция, непрерывная на связном множестве, принимает все значения из промежутка между двумя любыми своими значениями. Точнее:

если $a < b$, $f(a) = A$, $f(b) = B$ и C — некоторое число, лежащее между A и B, то найдется такая точка c, что $a < c < b$ и $f(c) = C$.

Особо выделяется частный случай этого утверждения, когда A и B имеют разные знаки:

если $f(a)$ и $f(b)$ имеют противоположные знаки, то в промежутке между a и b найдется такая точка c, что $f(c) = 0$.

Если область определения не является связным множеством, то это утверждение, разумеется, неверно, т. к., например, $\operatorname{tg}\frac{\pi}{4} = 1$, $\operatorname{tg}\left(3\frac{\pi}{4}\right) = -1$, но тем не менее ни в одной точке промежутка между $\frac{\pi}{4}$ и $3\frac{\pi}{4}$ функция $\operatorname{tg} x$ не обращается в 0.

О ПРЕДЕЛЕ ПОСЛЕДОВАТЕЛЬНОСТИ НЕПРЕРЫВНЫХ ФУНКЦИЙ

Рассмотрим на интервале $[0, 1]$ последовательность функций $f_n(x) = x^n$. Все члены последовательности — непрерывные функции, последовательность сходится, и ее предел

$$f(x) = \begin{cases} 0, & \text{если} \quad x < 1 \\ 1, & \phantom{\text{если}} \quad x = 1. \end{cases}$$

Эта функция имеет разрыв в точке 1. Причина заключается в том, что последовательность $f_n(x) = x^n$ сходится к $f(x)$ не равномерно. Это означает, что последовательность сходится в каждой точке, т. е. если $\varepsilon > 0$, то для каждой точки a интервала $[0, 1]$ найдется такое число $N(\varepsilon)$, что $| f_n(a) - f(a) | < \varepsilon$ лишь только $n > N(\varepsilon)$, но это $N(\varepsilon)$ зависит также и от a, и не удается для каждого $\varepsilon > 0$ найти такое $N(\varepsilon)$, которое годилось бы одновременно для всех точек a интервала $[0, 1]$. Пусть, например, $\varepsilon = \frac{1}{2}$. Тогда, как бы велико ни было число n, если a находится в достаточной близости от 1, то $| a^n - f(a) | = a^n > \frac{1}{2}$, так как при любом n $\lim\limits_{a \to 1} a^n = 1$.

Говорят, что *последовательность функций* $\{f_n(x)\}$ *равномерно сходится к функции* $f(x)$ *на множестве* H, если для любого числа $\varepsilon > 0$ найдется такое число $N(\varepsilon)$, что $|f_n(x) - f(x)| < \varepsilon$ лишь только $n > N(\varepsilon)$ для всех $x \in H$ (то есть N зависит лишь от ε и не зависит от x). Условия, при которых предел последовательности непрерывных функций есть функция непрерывная, сформулированы в следующей теореме:

если последовательность функций $\{f_n(x)\}$ равномерно сходится к функции $f(x)$ на множестве H, то функция $f(x)$ непрерывна в тех точках, в которых непрерывны все члены последовательности $\{f_n(x)\}$.

Поскольку сходимость ряда функций равносильна сходимости последовательности его частичных сумм, понятие равномерной сходимости можно распространить и на ряды функций. В этом случае можно сформулировать теорему, аналогичную предыдущей: если ряд, состоящий из непрерывных функций, сходится равномерно, то его сумма является непрерывной функцией.

ДИФФЕРЕНЦИАЛЬНОЕ ИСЧИСЛЕНИЕ

ПОНЯТИЕ ПРОИЗВОДНОЙ

1. Рассмотрим некоторое прямолинейное движение. Представим себе, что мы движемся по прямолинейной дороге, вдоль которой вместо километровых столбов проходит числовая ось. В нашем распоряжении имеется секундомер. С помощью таких приспособлений мы в любой момент времени t, показанный секундомером, можем определить пройденный путь s. Таким образом каждому значению t соответствует определенное значение s, то есть s является функцией t, что выражается с помощью уравнения движения:

$$s = s(t)\,.$$

График функции $s = s(t)$ называют *графиком пути*. Возникает вопрос, что мы понимаем под скоростью нашего движения в некоторый момент времени t_0. Заметим величину $s(t_0)$; через некоторое время, когда стрелка секундомера продвинется по шкале на Δt, то есть в момент времени $t_0 + \Delta t$, снова заметим пройденный путь. Тогда отрезок пути, пройденный за промежуток времени Δt,

$$s = s(t_0 + \Delta t) - s(t_0)\,.$$

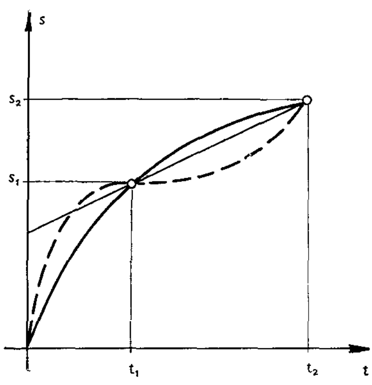
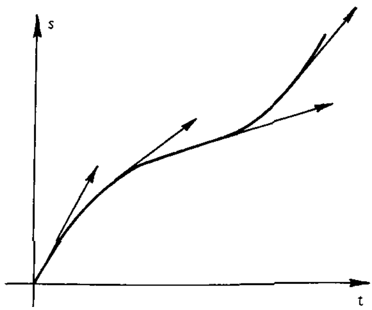


График пути. Пройденный путь есть функция времени

Отсюда после деления на Δt получаем: $\dfrac{\Delta s}{\Delta t} = \dfrac{s(t_0 + \Delta t) - s(t_0)}{\Delta t}$ (через Δ вооб-



Величина скорости равна угловому коэффициенту касательной к кривой

ще принято обозначать приращение, то есть величину изменения переменной). Полученное отношение приращений даст нам, очевидно, среднюю скорость движения за время Δt. Однако эту скорость ни в коем случае нельзя назвать средней скоростью движения в момент времени t_0, так как выбирая разные Δt, мы будем получать различные отношения приращений. Это становится еще заметнее, если с помощью графика пути установить геометрический смысл отношения приращений. На рисунке хорошо видно, что отношение приращений выражает угловой коэффициент секущей, проходящей через две соответствующие точки графика пути. Описанным выше способом, разумеется, нельзя определить мгновенную скорость движения, однако, математическое определение мгновенной скорости очевидно:

$$\lim_{\Delta t \to 0} \frac{s(t_0 + \Delta t) - s(t_0)}{\Delta t}.$$

Как уже упоминалось, это число нельзя получить в результате измерений, однако его связь с результатами измерений очевидна:

$$v(t_0) = \lim_{\Delta t \to 0} \frac{s(t_0 + \Delta t) - s(t_0)}{\Delta t}$$

является числом, от которого измеряемая средняя скорость отличается сколь угодно мало, если выбрать достаточно малое Δt.

Очевиден и геометрический смысл записанных выше пределов: точка $(t_2;\ s_2)$ секущей неограниченно приближается к точке $(t_1;\ s_1)$, то есть секущая стремится занять положение касательной, при этом угловой коэффициент секущей стремится к угловому коэффициенту касательной.

Таким образом, в каждой точке графика пути определена мгновенная скорость, величина которой равна угловому коэффициенту касательной в той же точке. Предел $\lim\limits_{\Delta t \to 0} \dfrac{s(t_0 + \Delta t) - s(t_0)}{\Delta t}$, то есть предел отношения приращений, называется производной функции $s(t)$ по переменной t в точке t_0, и обозначается через $\left(\dfrac{ds}{dt}\right)_{t=t_0}$ $\left(\text{здесь } \dfrac{ds}{dt} \text{ — единый символ, черта дроби не означает деления}\right)$.

2. Пусть $Q(t)$ — количество вещества, образовавшегося к моменту времени t в результате некоторой химической реакции. Под средней скоростью реакции между моментами времени t_0 и $t_0 + \Delta t$ понимают отношение количества вещества, образовавшегося за этот промежуток времени, к длине

самого промежутка:

$$\frac{Q(t_0+\Delta t)-Q(t_0)}{\Delta t}=\frac{\Delta Q}{\Delta t}.$$

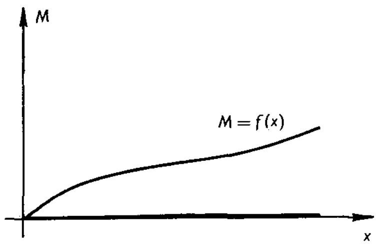
Мгновенную скорость реакции даст нам, очевидно, предел

$$\lim_{\Delta t\to 0}\frac{Q(t_0+\Delta t)-Q(t_0)}{\Delta t},$$

то есть производная от Q по t. Геометрически это значение производной выражает угловой коэффициент касательной к кривой $Q(t)$ в точке t_0.

3. Что следует понимать под плотностью неоднородного стержня в произвольной его точке? Выберем систему координат так, чтобы ось Ox проходила вдоль стержня, а точка O совпадала с его левым концом. Пусть функция $M(x)$ выражает массу части стержня от точки O до точки x. Тогда среднюю плотность части стержня от точки x_0 до точки $x_0+\Delta x$ можно определить формулой

$$\frac{M(x_0+\Delta x)-M(x_0)}{\Delta x}.$$



Плотность есть производная массы по x

Плотность стержня в точке x_0 выразится значением производной от M по x в этой точке:

$$\lim_{\Delta x\to 0}\frac{M(x_0+\Delta x)-M(x_0)}{\Delta x}=\left(\frac{dM}{dx}\right)_{x=x_0}.$$

Это значение производной выражает угловой коэффициент касательной к кривой $M(x)$ в точке x_0.

Из приведенных примеров видно, как нужно действовать, если мы хотим определить «крутизну» кривой, соответствующей некоторой функции $y=f(x)$, то есть угловой коэффициент касательной к кривой в некоторой точке $(x_0;\ f(x_0))$.

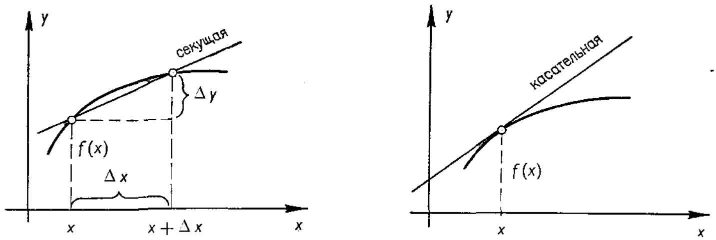
Угловой коэффициент касательной является пределом углового коэффициента секущей, угловой же коэффициент секущей дает нам отношение приращений:

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

Значит, угловой коэффициент касательной есть

$$\lim_{\Delta x\to 0}\frac{f(x_0+\Delta x)-f(x_0)}{\Delta x}=\left(\frac{dy}{dx}\right)_{x=x_0}=y'_{x=x_0},$$

то есть значение производной от переменной y по переменной x в точке x_0.



Определение углового коэффициента касательной

Можно указать еще много примеров, приводящих к подобным задачам нахождения пределов. Однако приведенные примеры вовсе не означают, что производная — это понятие физическое или, скажем, геометрическое. Речь идет просто о том, что мы имеем дело с пределом такого типа, который может быть использован для решения многочисленных задач физики, геометрии и других наук. Именно поэтому понятие производной относится к математическому анализу.

ПРАВИЛА ВЫЧИСЛЕНИЯ ПРОИЗВОДНЫХ

Дифференцирование некоторой функции $y = f(x)$, то есть вычисление ее производной, всегда сопряжено с нахождением некоторого предела.

Пример: если $y = x^2$, то

$$y' = \frac{dy}{dx} = \lim_{\Delta x \to 0} \frac{(x+\Delta x)^2 - x^2}{\Delta x} = \lim_{\Delta x \to 0} \frac{2x\,\Delta x + \Delta x^2}{\Delta x} = \lim_{\Delta x \to 0} (2x + \Delta x) = 2x.$$

Производные элементарных функций можно вычислять без труда, зная несколько важных пределов, подобных приведенному в примере, — так называемые правила вычисления производных. Приведем наиболее часто употребляемые из них.

1. Дифференцирование постоянной функции:

$$\boxed{y = c; \qquad y' = 0.}$$

2. Дифференцирование степенной функции:

$$\boxed{y = x^n; \qquad y' = nx^{n-1}.}$$

3. Дифференцирование суммы функций и функции, умноженной на постоянную:

$$\boxed{y = f_1(x) + f_2(x); \quad y' = f_1'(x) + f_2'(x); \quad y = C \cdot f(x); \quad y' = C \cdot f'(x).}$$

Пример: используя перечисленные до сих пор правила, мы уже можем дифференцировать многочлены.

Если $y = x^2 - 4x + 5$, до $y' = 2x - 4$.

4. Дифференцирование произведения:

$$\boxed{y = f_1(x) \cdot f_2(x); \qquad y' = f_1'(x) \cdot f_2(x) + f_1(x) \cdot f_2'(x).}$$

5. Дифференцирование отношения:

$$\boxed{y = \frac{f(x)}{g(x)}; \qquad y' = \frac{f'(x) \cdot g(x) - g'(x) \cdot f(x)}{[g(x)]^2}.}$$

Пример:

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

6. Дифференцирование тригонометрических функций:

$$\boxed{y = \sin x, \qquad y' = \cos x; \qquad y = \cos x, \qquad y' = -\sin x.}$$

7. Дифференцирование логарифмической функции:

$$\boxed{y = \ln x, \qquad y' = \frac{1}{x}; \qquad y = \log_a x, \qquad y' = \frac{1}{x \ln a}.}$$

Пример:

$$y = \sin x \cdot \ln x, \qquad y' = \cos x \cdot \ln x + \frac{\sin x}{x}.$$

8. Правило дифференцирования обратной функции:

$$\boxed{\frac{dy}{dx} = \frac{1}{\dfrac{dx}{dy}}.}$$

Поясним это правило на примерах:

а) Если $y = e^x$ (тогда $x = \ln y$), значит,

$$\frac{dy}{dx} = \frac{1}{\frac{dx}{dy}} = \frac{1}{\frac{1}{y}} = y, \qquad \frac{dx}{dy} = \frac{1}{y}, \qquad \frac{dy}{dx} = e^x,$$

б) Если $y = \sqrt{x}$, (тогда $x = y^2$), значит,

$$\frac{dy}{dx} = \frac{1}{\frac{dx}{dy}} = \frac{1}{2y}, \qquad \frac{dx}{dy} = 2y, \qquad \frac{dy}{dx} = \frac{1}{2} x^{-\frac{1}{2}}.$$

9. Дифференцирование гиперболических функций:

$$\boxed{y = \operatorname{sh} x, \quad y' = \operatorname{ch} x; \quad y = \operatorname{ch} x, \quad y' = \operatorname{sh} x; \quad y = \operatorname{tgh} x, \quad y' = \frac{1}{\operatorname{ch}^2 x}.}$$

10. Правило дифференцирования сложной функции:

$$\boxed{\left.\begin{aligned} y &= f(u) \\ u &= \varphi(x) \end{aligned}\right\} y = f[\varphi(x)] \quad \frac{dy}{dx} = \frac{dy}{du} \cdot \frac{du}{dx}.}$$

Примеры.

а)
$$y = \sin(x^2 + x + 1)$$
$$y = \sin u, \qquad u = x^2 + x + 1$$
$$y' = \frac{dy}{dx} = \frac{dy}{du} \cdot \frac{du}{dx} = [\cos(x^2 + x + 1)]\,(2x + 1)$$

б)
$$y = e^{\sin x}$$
$$y = e^u, \qquad u = \sin x$$
$$y' = \frac{dy}{dx} = \frac{dy}{du} \cdot \frac{du}{dx} = e^{\sin x} \cdot \cos x.$$

Правила дифференцирования обратной и сложной функций легко запоминаются, если заметить, что довольно часто с выражением $\frac{dy}{dx}$ можно производить вычисления как с обыкновенной дробью. В этом большое преимущество этого обозначения производной, введенного Лейбницем. Причина же такой возможности заключается в том, что $\frac{dy}{dx}$, хоть и не является отношением, но представляет собой предел отношения $\frac{\Delta y}{\Delta x}$.

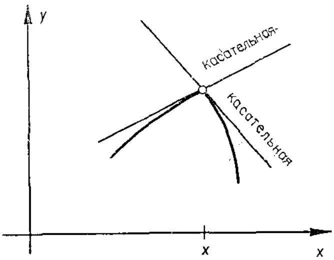
ДИФФЕРЕНЦИРУЕМОСТЬ И НЕПРЕРЫВНОСТЬ

Производная есть предел отношения приращений. Если в некоторой точке этот предел не существует, что говорят, то в этой точке функция *не дифференцируема*. Если функция в некоторой точке *дифференцируема*, то к кривой, изображающей функцию, можно построить (единственную) касательную в этой точке. Поэтому в точках, где кривая имеет изломы, функция, оставаясь непрерывной, не является дифференцируемой. На рисунке изображен график функции, для которого существуют лишь односторонние касательные, составляющие угол. Это означает, что пределы

$$\lim_{\Delta x \to x-0} \frac{f(x+\Delta x)-f(x)}{\Delta x}$$

и

$$\lim_{\Delta x \to x+0} \frac{f(x+\Delta x)-f(x)}{\Delta x}$$



Функция имеет производные слева и справа, но они не равны между собой

существуют, то есть отношение приращений имеет пределы справа и слева, называемые, соответственно, производной функции справа и слева, но эти пределы не равны между собой.

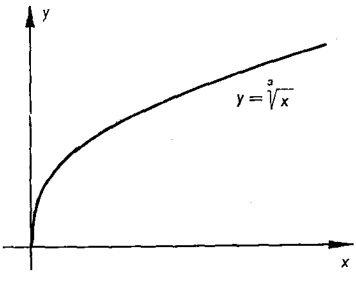
В качестве другого примера рассмотрим уже знакомую нам функцию

$$y=\begin{cases} x\cdot\sin\frac{1}{x}, & \text{если} \quad x\neq 0 \\ 0 & \qquad\quad\; x=0. \end{cases}$$

График этой функции (см. рис. на стр. 430) при стремлении x к 0 совершает бесконечное множество колебаний, но их амплитуда постоянно убывает. Как мы уже знаем, функция непрерывна в точке 0. Рассмотрим для нее отношение приращений в этой точке:

$$\frac{\Delta y}{\Delta x}=\frac{(0+\Delta x)\cdot\sin\frac{1}{0+\Delta x}-0}{\Delta x},$$

$$\frac{\Delta y}{\Delta x}=\frac{\Delta x\cdot\sin\frac{1}{\Delta x}}{\Delta x}=\sin\frac{1}{\Delta x}.$$



Функция $y=\sqrt[3]{x}$ не дифференцируема в точке 0

$\sin\frac{1}{\Delta x}$ при $\Delta x \to 0$ не имеет предела (даже одностороннего), следовательно, функция не дифференцируема в точке 0.

Функция $y=\sqrt[3]{\Delta x}$ не дифференцируема в точке 0, т. к.

$$\lim_{\Delta x\to 0}\frac{\sqrt[3]{\Delta x}}{\Delta x}=\lim_{\Delta x\to 0}\frac{1}{(\Delta x)^{2/3}}=\infty.$$

Более сложными методами можно построить пример и такой функции, которая непрерывна во всех точках некоторого интервала, но не дифференцируема ни в одной точке.

Дифференцируемость функции означает, что

$$\lim_{\Delta x\to 0}\frac{\Delta y}{\Delta x} \text{ существует (и конечен).}$$

Отсюда замечаем, что предел знаменателя равен 0. Но для существования конечного предела дроби необходимо, чтобы и предел числителя был равен 0:

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

а это означает, что функция непрерывна.

Из сказанного следует, что *непрерывность является необходимым, но не достаточным условием дифференцируемости.*

ДИФФЕРЕНЦИАЛ

В специальной научной и технической литературе очень часто встречается понятие дифференциала, поэтому целесообразно ознакомиться с этим понятием. Согласно определению производной

$$\frac{dy}{dx}=\lim_{\Delta x\to 0}\frac{\Delta y}{\Delta x}=\lim_{\Delta x\to 0}\frac{f(x+\Delta x)-f(x)}{\Delta x}$$

Если Δx достаточно мало, то

$$\frac{f(x+\Delta x)-f(x)}{\Delta x}=\frac{\Delta y}{\Delta x}$$

сколь угодно мало отличается от производной $\frac{dy}{dx}$, то есть «приближенно равно» ей:

$$\frac{f(x+\Delta x)-f(x)}{\Delta x}\approx f'(x).$$

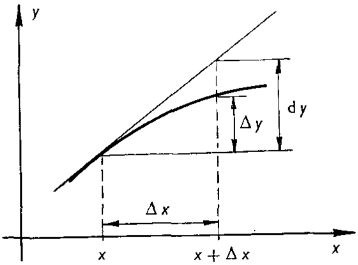
Отсюда

$$\Delta y = f(x+\Delta x)-f(x)\approx f'(x)\,\Delta x.$$

Приращение Δx называют дифференциалом переменной x и обозначают через dx. Приближенное значение приращения Δy называют дифференциалом функции y и обозначают через dy

Итак

$$\boxed{dy = f'(x)\,dx.}$$



Если dx достаточно мало, то dy с любой степенью точности приближает приращение функции

Как видно из формул, дифференциал функции $y = f(x)$ зависит и от x, и от dx. На рисунке хорошо видно, что дифференциал dy сколь угодно мало отличается от приращения функции Δy, если dx достаточно мало. В самом деле, в то время как Δy есть приращение ординаты кривой, dy является соответственным приращением ординаты касательной. Между тем, если в достаточно малой окрестности некоторой точки (x, y) вместо кривой рассмотреть касательную к ней в этой точке, то возникающая при этом погрешность сколь угодно мала, больше того, она пренебрежимо мала даже в сравнении с величинами dx и dy. Это означает, что

$$\lim_{dx\to 0}\frac{dy-\Delta y}{dx}=0.$$

ЗНАЧЕНИЕ ЗНАКА ПРОИЗВОДНОЙ

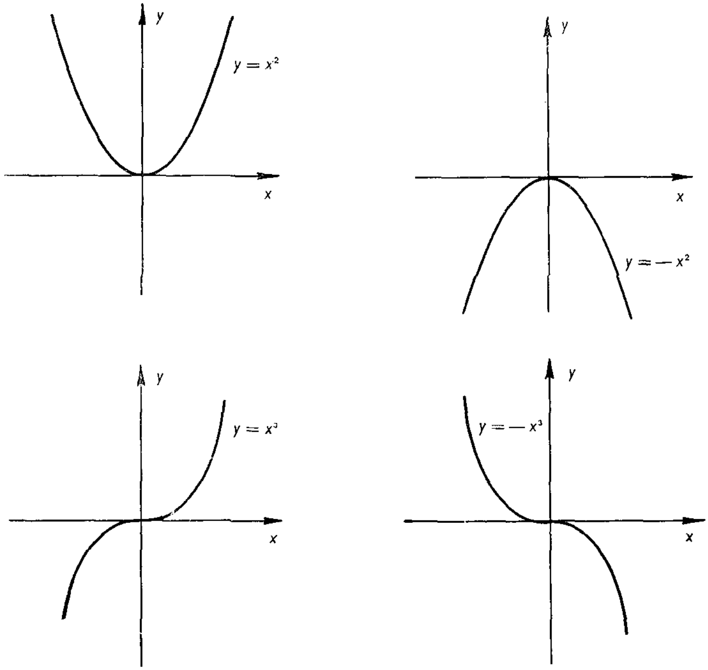
Выше уже шла речь о том, что означает, если функция $y = f(x)$ монотонно возрастает на некотором множестве. Теперь мы рассмотрим это понятие с другой стороны.

Функция $y = f(x)$ называется монотонно возрастающей в точке x_0, если существует такая окрестность точки x_0, для всех точек x которой из $x < x_0$ $(x_0 < x)$ следует $f(x) < f(x_0)$ $[f(x_0) < f(x)]$. Аналогично определяется понятие монотонно убывающей функции.

Зная, какой знак имеет производная, можно сделать важное заключение о ходе изменения функции:

если в некоторой точке x $f'(x) > 0$, то функция $f(x)$ в этой точке монотонно возрастает, если же $f'(x) < 0$, то убывает.

Если $f'(x) = 0$, то о поведении функции ничего нельзя сказать. Действительно, рассмотрим функции $y = x^2$, $y = -x^2$, $y = x^3$, $y = -x^3$. Их производные равны, соответственно, $2x$, $-2x$, $3x^2$, $-3x^2$. В точке $x = 0$ все эти производные обращаются в 0, в то время как в этой точке первая из перечисленных функций имеет минимум, вторая — максимум, третья монотонно возрастает, четвертая — убывает.

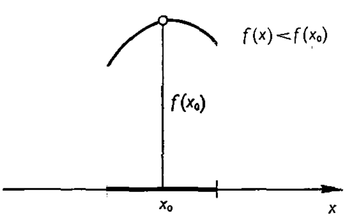


Из равенства нулю первой производной еще нельзя сделать заключения о поведении функции

Говорят, что функция $y = f(x)$ имеет в точке x_0 *локальный максимум* если существует такая окрестность точки x_0, для всех точек x которой $(x \neq x_0)$ выполняется неравенство $f(x) < f(x_0)$. Аналогично определяется *локальный минимум* функции.

Иногда — если это не приводит к недоразумению — слово «локальный» опускают. Для обозначения максимума и минимума существует объединяющий их термин — экстремум.

Для того, чтобы функция $f(x)$ в точке x_0 имела экстремум, необходимо, чтобы $f'(x_0) = 0$.



Функция $f(x)$ имеет локальный максимум в точке $x = x_0$

Это очевидно, поскольку, если функция в точке x_0 имеет экстремум, то в этой точке она не может быть ни возрастающей, ни убывающей, то есть ее производная в точке x_0 не может быть ни положительной, ни отрицательной. Но следует отметить, что это верно лишь для дифференцируемой функции, например $y = |x|$ имеет экстремум (минимум) при $x = 0$, хотя $y' \neq 0$, поскольку не существует в этой точке.

Следовательно, для существования экстремума дифференцируемой функции необходимо, чтобы $f'(x) = 0$.

Однако, как показывает пример функции $y = x^3$, это условие не является достаточным, так как хотя производная в точке $x = 0$ и равна 0, функция в этой точке возрастает.

Если x_0 — точка максимума, то слева от x_0 функция возрастает, а справа — убывает. Так будет в том случае, если $f'(x_0) = 0$, и в достаточно малой окрестности точки x_0 слева от нее $f'(x) > 0$, а справа $f'(x) < 0$. С помощью аналогичного рассуждения получаем, что функция имеет минимум в точке x_0, если $f'(x_0) = 0$, и в достаточно малой окрестности точки x_0 слева от нее $f'(x) < 0$, а справа $f'(x) > 0$.

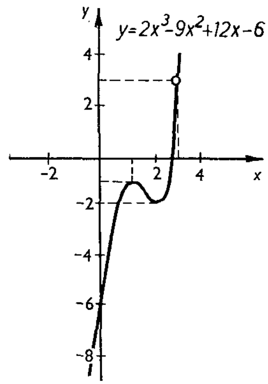
Подведем итог: функция $f(x)$ имеет в точке x_0 экстремум, если $f'(x_0) = 0$, и в точке x_0 производная функции меняет знак; если она меняет знак с отрицательного на положительный, то в точке x_0 функция имеет минимум, если же с положительного на отрицательный — максимум.

Пример:

$$y = 2x^3 - 9x^2 + 12x - 6,$$

$$y' = 6x^2 - 18x + 12 = 6(x - 1)(x - 2).$$

$y'(1) = y'(2) = 0$, и y' в точке $x = 1$ меняет знак с положительного на отрицательный, а в точке $x = 2$ — с отрицательного на положительный. Значит, в точке $x = 1$ функция имеет максимум, а в точке $x = 2$ — минимум.



В точке $x = 1$ функция имеет максимум, в точке $x = 2$ — минимум

ЗНАЧЕНИЕ ЗНАКА ВТОРОЙ ПРОИЗВОДНОЙ ПРИ ОТЫСКАНИИ ЭКСТРЕМУМОВ

Вторая производная функции — это производная от ее производной. Для ее обозначения используют символ $\frac{d^2y}{dx^2}$ или y''.

Аналогично вводится понятие третьей, четвертой и вообще производной n-го порядка.

На практике при решении конкретных задач применение достаточного условия существования экстремума, с которым мы познакомились выше, часто вызывает большие затруднения. Действительно, не так уж просто установить, что производная равна 0 и при этом меняет знак. Целесообразно поэтому исследовать поведение второй производной. Приняв во внимание тот факт, что если $f'(x)$, монотонно возрастая, обращается в 0, то неизбежно меняет знак, можно сформулировать следующую теорему, дающую нам значительно более простое условие экстремума:

Если $f'(x_0) = 0$ и $f''(x_0) < 0$, то функция $f(x)$ в точке x_0 имеет максимум; если же $f'(x_0) = 0$ и $f''(x_0) > 0$, то x_0 — точка минимума функции.

Примеры:

1. Какие размеры должна иметь литровая кастрюля, если на ее изготовление требуется использовать как можно меньше листового металла? (Кастрюля должна иметь форму цилиндра без крышки, количеством металла, используемого в местах спайки, можно пренебречь).

$$F = r^2\pi + 2r\pi \cdot m; \qquad r^2\pi \cdot m = 1; \qquad m = \frac{1}{r^2\pi},$$

$$F = r^2\pi + \frac{2}{r},$$

$$F' = 2r\pi - \frac{2}{2},$$

$$F'' = 2\pi + \frac{4}{r^3}.$$

Решением уравнения $F' = 0$ является $r = \frac{1}{\sqrt[3]{\pi}}$. Поскольку $F''\left(\frac{1}{\sqrt[3]{\pi}}\right) >$ > 0, то кастрюля с наименьшей поверхностью должна иметь следующие размеры:

$$r = \frac{1}{\sqrt[3]{\pi}}, \qquad m = \frac{1}{\sqrt[3]{\pi^2}}.$$

2. Наиболее подходящее значение результата экспериментов.

Как бы ни были точны приборы, абсолютно точных измерений произвести нельзя. Для получения возможно более точного результата эксперимент (измерение) производят несколько раз. В результате n измерений получают n данных, как правило, отличающихся друг от друга. Обозначим их, соответственно, через

$$a_1, a_2, \ldots, a_n.$$

На основании этих данных требуется установить возможно более правильный результат. Обозначим его через x.

Наиболее правильным естественно назвать такое число x, которое по возможности меньше отличается от данных измерений. Величины ошибок, допускаемых при отдельных измерениях, есть

$$\pm|x - a_1|, \pm|x - a_2|, \ldots, \pm|x - a_n|,$$

со знаком + или — согласно тому, превышает ли результат измерения величину x или нет. Нас, однако, интересует величина ошибки, а не ее знак, поэтому лучше вместо величин ошибок рассматривать их квадраты, которые всегда положительны. Итак, наиболее подходящим значением x мы будем считать такое, для которого сумма квадратов ошибок, отклонений от него, допущенных при отдельных измерениях, минимальна. Таким образом задача свелась к следующей: при каком значении x функция

$$y = (x - a_1)^2 + (x - a_2)^2 + \dots + (x - a_n)^2$$

имеет минимум?

Производная функции должна равняться 0, то есть

$$y' = 2(x - a_1) + 2(x - a_2) + \dots + 2(x - a_n) = 0,$$

или

$$nx - (a_1 + a_2 \dots + a_n) = 0,$$

отсюда

$$x = \frac{a_1 + a_2 + \dots + a_n}{n}.$$

Нетрудно видеть, что при этом значении x функция действительно имеет минимум. В самом деле,

$$y = 2 + 2 + \dots + 2 = 2n > 0.$$

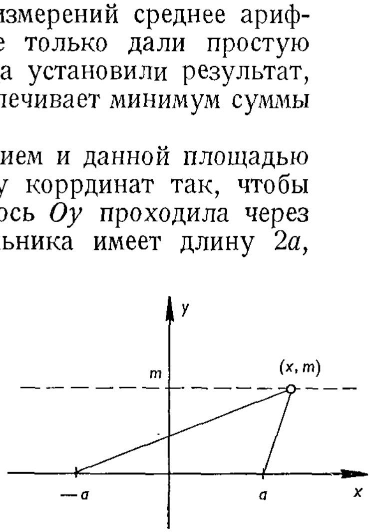
Таким образом, приняв в качестве результата измерений среднее арифметическое экспериментальных данных, мы не только дали простую формулу для его вычисления, но раз и навсегда установили результат, наиболее подходящий в том смысле, что он обеспечивает минимум суммы квадратов ошибок.

3. Какой из треугольников с данным основанием и данной площадью имеет наименьший периметр? Выберем систему коррдинат так, чтобы основание треугольника лежало на оси Ox, а ось Oy проходила через середину основания. Пусть основание треугольника имеет длину $2a$, тогда его концы расположены в точках $(-a;\ 0)$ и $(a;\ 0)$. Если высота треугольника равна m, то его вершина лежит на прямой $y = m$ и имеет координаты $(x;\ m)$, где x — переменная. Периметр треугольника задается функцией

$$f(x) = \sqrt{(x+a)^2 + m^2} + \sqrt{(x-a)^2 + m^2} + 2a$$

$$f'(x) = \frac{x+a}{\sqrt{(x+a)^2 + m^2}} + \frac{x-a}{\sqrt{(x-a)^2 + m^2}}$$

$$f''(x) = \frac{m^2}{\left(\sqrt{(x+a)^2 + m^2}\right)^3} + \frac{m^2}{\left(\sqrt{(x-a)^2 + m^2}\right)^3}.$$



Из всех треугольников с данным основанием и данной площадью наименьший периметр имеет равнобедренный треугольник

Нетрудно видеть, что, с одной стороны, $f'(0) = 0$, с другой, $f''(x)$ всюду положительна. Отсюда следует, что $f'(x)$ монотонно возрастает и, следовательно, обращается в нуль лишь в одной точке. В этой точке функция $f(x)$ имеет минимум. Это значит, что решением задачи, то есть треугольником с наименьшим периметром при данном основании и данной площади, является равнобедренный треугольник.

Этот пример, если попытаться решить его другим способом, оказывается весьма поучительным. Он иллюстрирует тот факт, что решениями задач на отыскание экстремумов не всегда являются локальные экстремумы, а потому не могут быть найдены путем отыскания точек, в которых $f'(x) = 0$. Из элементарной геометрии нам известна формула $t = \varrho s$, где t — площадь треугольника, ϱ — радиус вписанной в него окружности, а s — половина периметра. Если исходить из этой формулы, то следует рассмотреть функцию

$$s = \frac{t}{\varrho} = f(\varrho).$$

Однако производная

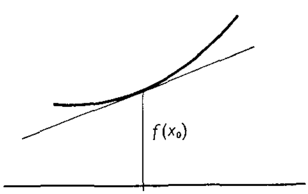
$$\frac{ds}{d\varrho} = -\frac{t}{\varrho^2}$$

нигде не обращается в нуль, будучи всюду отрицательной. Значит, функция $f(\varrho)$ монотонно убывает и, следовательно, для произвольно взятого интервала она достигает своего минимума в правом конце интервала. Однако при данной площади t треугольника радиус ϱ вписанной в него окружности не может быть произвольно большим, поэтому решение задачи в этом случае дает нам $\sup \varrho$.

ВЫПУКЛЫЕ И ВОГНУТЫЕ ФУНКЦИИ, ТОЧКИ ПЕРЕГИБА

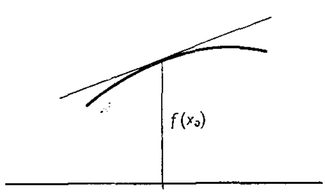
Функция $y = f(x)$ называется *выпуклой* (вниз) в точке x_0, если в достаточно малой окрестности точки x_0 график функции расположен над касательной к ней в точке $(x_0; f(x_0))$ и *вогнутой* (выпуклой вверх) в противном случае.

Условия, определяющие форму графика функции, приведены в следующей теореме:



a)

Выпуклая кривая

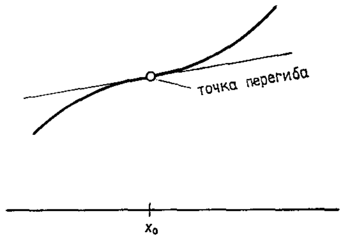


b)

Вогнутая кривая

если в точке x_0 $y'' > 0$, *то в этой точке функция выпукла; если в некоторой точке* x_1 $y'' < 0$, *то в такой точке функция вогнута.*

Точки, в которых функция (кривая) из выпуклой переходит в вогнутую или наоборот, называются *точками перегиба*. В точке перегиба касательная пересекает кривую. Необходимое условие точки перегиба: $y'' = 0$; достаточное условие: $y'' = 0$ и при этом меняет знак. Таким образом отыскание точек перегиба функции означает нахождение экстремумов ее производной.



В точке перегиба касательная пересекает кривую

Пример: Найти точки перегиба функции (см. рисунок на стр. 433)

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

$$y' = \frac{1 - x^2}{(1 + x^2)^2}, \quad y'' = -2x \frac{3 - x^2}{(1+x^2)^2}$$

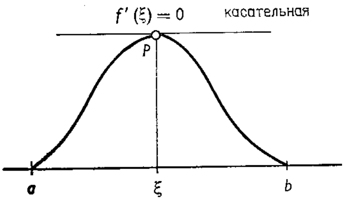
$y'' = 0$, если $x = -\sqrt{3}$, 0, $\sqrt{3}$, и во всех трех точках y'' меняет знак.

ЗАМЕЧАНИЯ О ЛОКАЛЬНЫХ И ГЛОБАЛЬНЫХ (ОБЩИХ) СВОЙСТВАХ

Мы назвали функцию $y = f(x)$ монотонно возрастающей на множестве H, если для любых двух точек множества H из $x_1 < x_2$ следует $f(x_1) < f(x_2)$. Было сформулировано и другое определение: функция называется монотонно возрастающей в точке x_0, если существует такая окрестность точки x_0, для всех точек x которой из $x < x_0$ $(x_0 < x)$ следует $f(x) < f(x_0)$ $[f(x_0) < f(x)]$. Первое определение устанавливает понятие монотонности как *глобальное* свойство, имеющее место на множестве точек, в то время как второе определение трактует монотонность как локальное свойство, выполняющееся в одной точке (и в некоторой ее окрестности). Оба определения, естественно, согласуются друг с другом: если функция монотонно возрастает в глобальном смысле, то она возрастает и локально. С другой стороны, применяя теорему о среднем значении, можно показать, что если функция локально возрастает в каждой точке некоторого интервала, то в этом интервале она возрастает и в глобальном смысле. Выпуклость и вогнутость функций обычно также определяют в глобальном смысле: функция называется выпуклой в интервале $(a; b)$, если любая дуга ее графика расположена под стягивающей ее хордой, и вогнутой, если любая дуга расположена над стягивающей ее хордой.

ТЕОРЕМЫ О СРЕДНЕМ ЗНАЧЕНИИ

Теорема Ролля: если функция $f(x)$ дифференцируема в замкнутом интервале $[a;\ b]$, и $f(a) = f(b) = 0$, то внутри интервала найдется такая точка ξ, что $f'(\xi) = 0$. Такой точкой будет, очевидно, точка минимума или максимума функции. Геометрический смысл теоремы Ролля следующий: если график функции таков, что можно провести хорду, параллельную оси Ox, то на соответствующей дуге графика найдется точка P, где касателньая параллельна оси Ox, а значит, и хорде.

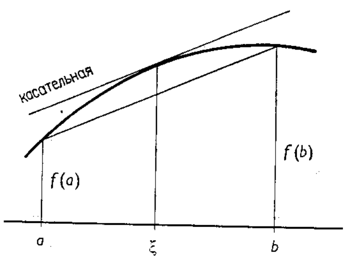


В точке P касательная параллельна оси x

Если, исходя из этой формулировки, обобщить теорему на случай произвольной хорды, то мы придем к *теореме Лагранжа о среднем значении*:

если функция $f(x)$ дифференцируема в замкнутом интервале $[a;\ b]$, то внутри интервала найдется такая точка ξ, что

$$f'(\xi) = \frac{f(b) - f(a)}{b - a}.$$



Теорема Лагранжа о среднем значении

Пусть $b - a = h$. Поскольку существует такое число ϑ $(0 < \vartheta < 1)$, что $\xi = a + \vartheta\xi$, то предыдущее равенство можно переписать так:

$$\frac{f(a + h) - f(a)}{h} = f'(a + \vartheta h)$$

$$0 < \vartheta < 1.$$

НЕСКОЛЬКО ВАЖНЫХ СЛЕДСТВИЙ ИЗ ТЕОРЕМЫ О СРЕДНЕМ ЗНАЧЕНИИ

1. Если во всех точках x интервала $[a;\ b]$ $f'(x) = 0$, то $f(x) = C$, то есть только *постоянная* обладает тем свойством, что ее производная тождественно равна 0.

2. Если во всех точках x интервала $[a;\ b]$ $f'(x) = g'(x)$, то функции $f(x)$ и $g(x)$ различаются лишь на аддитивную постоянную.

3. Если функция $f(x)$ дифференцируема в открытом интервале $(a;\ b)$, и ее производная внутри интервала ограничена, то ограничена и сама функция $f(x)$. (Пример функции $y = \lg x$, рассматриваемой в интервале $(0;\ 1)$, показывает, что если производная не ограничена, то и функция не обязательно ограничена.)

Теорему Лагранжа о среднем значении можно обобщить дальше. Идея обобщения следующая: рассмотрим равенство

$$\frac{f(b)-f(a)}{b-a}=f'(\xi).$$

Слева записано отношение величины приращения функции $f(x)$ к величине приращения переменной x (которая также может рассматриваться как функция: $g(x)=x$), справа — значение производной. Однако правую часть можно переписать иначе:

$$\frac{f'(\xi)}{1}.$$

Тогда исходное равенство примет вид:

$$\frac{f(b)-f(a)}{b-a}=\frac{f'(\xi)}{1}.$$

Теперь слева записано отношение величины приращения функции $f(x)$ к величине приращения функции $g(x)=x$, справа же — отношение значения производной функции $f(x)$ в точке ξ к значению производной функции $g(x)$ в той же точке (производная функции $g(x)=x$ тождественно равна 1). Идея обобщения после этого ясна. Таким образом, мы приходим к теореме Коши о среднем значении:

если функции $y=f(x)$ и $y=g(x)$ дифференцируемы в интервале $[a;\ b]$, и $g'(x)$ внутри интервала не обращается в нуль, то найдется такая точка ξ, лежащая между a и b, что

$$\frac{f(b)-f(a)}{g(b)-g(a)}=\frac{f'(\xi)}{g'(\xi)}.$$

Условие $g(x)\neq 0$ необходимо, чтобы избежать возможного нуля в знаменателе.

ПРАВИЛО ЛОПИТАЛЯ

Рассмотрим следующий предел:

$$\lim_{x\to 0}\frac{1-\cos x}{x^2}.$$

Поскольку и в числителе, и в знаменателе стоят непрерывные функции, предел каждого из них получается простой подстановкой. Однако в данном случае и предел числителя, и предел знаменателя равен 0, и нельзя применить теорему о пределе отношения, так как отношение пределов есть $\frac{0}{0}$.

Вообще в случае предела вида

$$\lim_{x \to a} \frac{f(x)}{g(x)},$$

когда

$$\lim_{x \to a} f(x) = f(a) = 0 \quad \text{и} \quad \lim_{x \to a} g(x) = g(a) = 0,$$

то есть и предел числителя, и предел знаменателя равен 0 (в результате подстановки получаются нули), мы имеем дело с так называемой неопределенностью типа

$$\frac{0}{0}.$$

В таких случаях применимо следующее правило:

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

Точный смысл записанной формулы следующий: если существует предел в правой части равенства, то существует и предел в левой части, при этом оба предела равны. Таким образом в случае неопределенности типа $\frac{0}{0}$ вместо предела

$$\lim_{x \to a} \frac{f(x)}{g(x)}$$

можно искать предел

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

Пример:

$$\lim_{x \to 0} \frac{1 - \cos x}{x^2},$$

здесь $f(x) = 1 - \cos x$; $g(x) = x^2$; $f(x) \to 0$, $g(x) \to 0$. Продифференцируем числитель и знаменатель. Получим:

$$\lim_{x \to 0} \frac{1 - \cos x}{x^2} = \lim_{x \to 0} \frac{\sin x}{2x} = \frac{1}{2} \lim_{x \to 0} \frac{\sin x}{x} = \frac{1}{2},$$

поскольку, как известно, $\lim\limits_{x \to 0} \frac{\sin x}{x} = 1$.

Аналогично можно действовать в случае неопределенности типа $\frac{\infty}{\infty}$.

Например,

$$\lim_{x \to \infty} \frac{x^2}{e^x} = \lim_{x \to \infty} \frac{2x}{e^x}.$$

Мы опять получили неопределенность типа $\frac{\infty}{\infty}$. Применим правило Лопиталя еще раз:

$$\lim_{x \to \infty} \frac{x^2}{e^x} = \lim_{x \to \infty} \frac{2x}{e^x} = \lim_{x \to \infty} \frac{2}{e^x} = 0.$$

Применяя правило Лопиталя n раз, можно точно так же получить, что

$$\lim_{x \to \infty} \frac{x^n}{e^x} = 0.$$

РЯД ТЕЙЛОРА

ДЛЯ ЧЕГО НУЖНА ФОРМУЛА ТЕЙЛОРА?

Из всех функций различные математические выкладки проще всего производить с *целыми рациональными функциями* или, как их еще называют, *многочленами* (см. главу «Алгебра»).

Рассмотрим, например, подстановку, то есть вычисление значения функции в некоторой точке. Возьмем к примеру многочлен

$$f(x) = 5x^4 - 3x^3 - 2x^2 + 8x - 7$$

и подставим в него $x = 1{,}2$:

$$f(1{,}2) = 4{,}904.$$

Это может вычислить всякий, кто знаком с четырьмя арифметическими действиями. Сложнее обстоит дело, если вместо многочлена рассмотреть какую-нибудь другую фукнцию. Пусть, например,

$$f(x) = e^x,\ f(1,2) = e^{1,2}.$$

Значение этой функции вычислить гораздо труднее, особенно тому, кто знает лишь четыре арифметических действия. Тем не менее, большинство функций, встречающихся на практике, именно таково. Например, значения логарифмической или тригонометрических функций вычисляют не подстановкой, а при помощи таблиц. Такие таблицы, однако, надо было изготовить, а для этого значения логарифмической и тригонометрических функций надо было каким-то образом вычислить.

Аналогично положение и в случае дифференцирования — вычислять производные многочленов очень легко. Как мы увидим в дальнейшем, то же самое наблюдается и при интегрировании.

Подводя итог, можно сказать, что иметь дело с многочленами проще, чем с другими функциями. Целесообразно поэтому рассмотреть следующую проблему: как можно с помощью многочленов приближать другие функции, встречающиеся на практике.

ФОРМУЛА ТЕЙЛОРА ДЛЯ МНОГОЧЛЕНОВ

Рассмотрим многочлен второй степени:

$$f(x) = a_0 + a_1 x + a_2 x^2,$$
$$f'(x) = a_1 + 2a_2 x,$$
$$f''(x) = 2a_2.$$

Подставляя во все три функции $x = 0$, получаем:

$$a_0 = f(0),\ a_1 = \frac{f'(0)}{1!},\ a_2 = \frac{f''(0)}{2!}.$$

Таким образом, многочлен второй степени может быть записан в следующей форме:

$$f(x) = f(0) + \frac{f'(0)}{1!} x + \frac{f''(0)}{2!} x^2.$$

Трижды дифференцируя многочлен третьей степени, аналогичным образом получаем:

$$f(x) = a_0 + a_1 x + a_2 x^2 + a_3 x^3,$$
$$f'(x) = a_1 + 2a_2 x + 3a_3 x^2,$$
$$f''(x) = 2a_2 + 3 \cdot 2a_3 x,$$
$$f'''(x) = 3 \cdot 2a_3.$$

Отсюда, подставляя $x = 0$, имеем:

$$a_0 = f(0),\ a_1 = \frac{f'(0)}{1!},\ a_2 = \frac{f''(0)}{2!},\ a_3 = \frac{f'''(0)}{3!}.$$

Вообще для многочлена n-ой степени:

$$f(x) = a_0 + a_1 x + a_2 x^2 + \ldots + a_n x^n,$$
$$a_0 = f(0),\ a_1 = \frac{f'(0)}{1!},\ a_2 = \frac{f''(0)}{2!}, \ldots a_n = \frac{f^{(n)}(0)}{n!}.$$

Эти формулы раскрывают важное и характерное свойство многочленов: если известны значения многочлена n-ой степени и первых его n производных в точке 0, то тем самым — поскольку известны его коэффициенты — многочлен определен однозначно.

Проиллюстрируем сказанное на примере.

Найти многочлен пятой степени, удовлетворяющий следующим условиям:

$$f(0) = 2,\ f'(0) = -1,\ f''(0) = 3,\ f'''(0) = 4,\ f^{(IV)}(0) = -2,\ f^{(V)}(0) = 1.$$

Решение:

$$f(x) = 2 - \frac{1}{1!}x + \frac{3}{2!}x^2 + \frac{4}{3!}x^3 - \frac{2}{4!}x^4 + \frac{1}{5!}x^5.$$

В предыдущих рассуждениях точка $x = 0$ играла особую роль. Естественно задать вопрос, останется ли верной сформулированная выше теорема, если вместо нуля взять произвольную точку a? Оказывается, что в этом случае теорема звучит так:

если в некоторой точке $x = a$ заданы значения многочлена n-ой степени и первых его n производных, то тем самым многочлен определен однозначно.

Но если до сих пор значения многочлена и его производных в точке 0 давали нам величины коэффициентов и таким образом определяли сам многочлен, то теперь положение несколько иное. Ведь даже постоянный член, который, как мы видели, равен $f(0)$, не может быть выражен через $f(a)$. Рассмотрим, однако, случай, когда многочлен разложен не по степеням x, а по степеням $(x - a)$. Возьмем к примеру многочлен четвертой степени:

$$\begin{aligned}
f(x) &= a_0 + a_1(x - a) + a_2(x - a)^2 + a_3(x - a)^3 + a_4(x - a)^4\\
f'(x) &= a_1 + 2a_2(x - a) + 3a_3(x - a)^2 + 4a_4(x - a)^3\\
f''(x) &= 2a_2 + 3 \cdot 2a_3(x - a) + 4 \cdot 3 \cdot a_4(x - a)^2\\
f'''(x) &= 3 \cdot 2a_3 + 4 \cdot 3 \cdot 2 \cdot a_4(x - a)\\
f^{(IV)}(x) &= 4 \cdot 3 \cdot 2 \cdot a_4.
\end{aligned}$$

Подставляя $x = a$, получаем:

$$a_0 = f(x),\ a_1 = \frac{f'(a)}{1!},\ a_2 = \frac{f''(a)}{2!},\ a_3 = \frac{f'''(a)}{3!},\ a_4 = \frac{f^{(IV)}(a)}{4!},$$

то есть значения функции и ее производных в точке a опять однозначно определяют коэффициенты многочлена, разница лишь в том, что многочлен в этом случае должен быть разложен не по степеням x, а по степеням $(x - a)$. Вообще для многочлена n-ой степени справедлива следующая формула:

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

Пример:

Найти многочлен третьей степени, удовлетворяющий следующим условиям:

$$f(2) = 3,\ f'(2) = -2,\ f''(2) = 3,\ f'''(2) = 1\,.$$

Решение:

$$f(x) = 3 - \frac{2}{1!}(x-2) + \frac{3}{2!}(x-2)^2 + \frac{1}{3!}(x-2)^3.$$

Нетрудно, разумеется, записать разложение этого многочлена и по степеням x. Для этого нужно лишь возвести $(x-2)$ в квадрат и в куб, умножить на соответствующие коэффициенты и привести подобные члены. В результате получим, что

$$f(x) = \frac{35}{3} - 6x + \frac{1}{2}x^2 + \frac{1}{6}x^3,$$

то есть однозначно определен многочлен третьей степени, удовлетворяющий условиям:

$$f(0) = \frac{35}{3},\ \frac{f'(0)}{1!} = -6,\ \frac{f''(0)}{2!} = \frac{1}{2},\ \frac{f'''(0)}{3!} = \frac{1}{6}\,.$$

Итак, если $f(x)$ — многочлен n-ой степени, то

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

Это и есть формула Тейлора для многочлена степени n.

Формула Маклорина. Если вместо a подставить 0, то мы получим частный случай формулы Тейлора, носящий название формулы Маклорина:

$$f(x) = f(0) + \frac{f'(0)}{1!}x + \frac{f''(0)}{2!}x^2 + \ldots + \frac{f^{(n)}(0)}{n!}x^n.$$

ФОРМУЛА ТЕЙЛОРА

Предположим, что функция $f(x)$ — как большинство функций, встречающихся на практике, — бесконечно дифференцируема (то есть для нее существуют производные любого порядка). В этом случае для функции можно записать так называемый *многочлен Тейлора n-го порядка*:

$$\boxed{\begin{array}{l} f(a)+\dfrac{f'(a)}{1!}(x-a)+\dfrac{f''(a)}{2!}(x-a)^2+\ldots+\dfrac{f^{(n)}(a)}{n!}(x-a)^n=T_n(x) \\ \text{и} \\ \quad T_n(a)=f(a)\,;\; T_n'(a)=f'(a)\,;\; T_n''(a)=f''(a)\,;\;\ldots\,;\; T_n^{(n)}(a)=f^{(n)}(a). \end{array}}$$

Следовательно, многочлен Тейлора обладает тем свойством, что он и первые его n производных в точке a принимают, соответственно, те же значения, что и функция $f(x)$ и ее производные. Поскольку теперь функция $f(x)$ не является многочленом n-ой степени, то равенство $f(x)=T_n(x)$ не может иметь места. Проверим, однако, не является ли многочлен $T_n(x)$ хорошим приближением для $f(x)$. С этой целью рассмотрим следующее равенство:

$$f(x)=f(a)+\frac{f'(a)}{1!}(x-a)+\frac{f''(a)}{2!}(x-a)^2+\ldots+\frac{f^{(n)}(a)}{n!}(x-a^n)+r_n\,,$$

где через r_n обозначена разность между $f(x)$ и $T_n(x)$. r_n обычно называют n-ым дополнительным членом или просто дополнительным членом.

Для r_n имеет место следующая формула:

$$\boxed{r_n=\frac{f^{(n+1)}(\xi)}{(n+1)!}(x-a)^{n+1}\,,}$$

где точка ξ лежит между a и x.

Подведем итог:

$$\boxed{\begin{array}{l} f(x)=f(a)+\dfrac{f'(a)}{1!}(x-a)+\dfrac{f''(a)}{2!}(x-a)^2+\ldots+\dfrac{f^{(n)}(a)}{n!}(x-a)^n+r_n \\ r_n=\dfrac{f^{(n+1)}(\xi)}{(n+1)!}(x-a)^{n+1}. \end{array}}$$

Это есть формула Тейлора. Полагая здесь $x=0$, получим формулу Маклорина:

$$\boxed{f(x)=f(0)+\frac{f'(0)}{1!}x+\frac{f''(0)}{2!}x^2+\ldots+\frac{f^{(n)}(0)}{n!}x^n+\frac{f^{(n+1)}(\xi)}{(n+1)!}x^{n+1}\,.}$$

Напишем разложение в ряд Маклорина функции $f(x) = e^x$:

$$f'(x) = e^x,\ f''(x) = e^x, \ldots, f^{(n)}(x) = e^x,\ f^{(n+1)}(x) = e^x,$$

$$f(0) = 1,\ f'(0) = 1,\ f''(0) = 1,\ \ldots,\ f^{(n)}(0) = 1,\ f^{(n+1)}(\xi) = e^\xi = e^{\vartheta x}.$$

Отсюда:

$$e^x = 1 + \frac{x}{1!} + \frac{x^2}{2!} + \ldots + \frac{x^n}{n!} + r_n.$$

$$r_n = \frac{f^{(n+1)}(\xi)}{(n+1)!} \cdot x^{n+1} = \frac{e^\xi}{(n+1)!} x^{n+1}.$$

Поскольку $0 < \xi < x$, то $\xi = 0 + \vartheta x = \vartheta x$ $(0 < \vartheta < 1)$ и, следовательно,

$$r_n = \frac{e^{\vartheta x}}{(n+1)!} x^{n+1}.$$

Используем полученный результат для приближенного вычисления числа e. Положим $x = 1$.

$$e = 1 + \frac{1}{1!} + \frac{1}{2!} + \ldots + \frac{1}{n!} + r_n,\quad r_n = \frac{e^\vartheta}{(n+1)!}.$$

В то же время из предыдущего нам известно, что

$$e = \lim_{n\to\infty} \left(1 + \frac{1}{n}\right)^{n+1},$$

где последовательность $\left(1 + \frac{1}{n}\right)^{n+1}$ монотонно убывает. Пусть, например, $n = 5$:

$$\left(1 + \frac{1}{5}\right)^6 = \left(\frac{6}{5}\right)^6 \frac{46656}{15625} < 3,$$

Следовательно, $e < 3$, но тогда и $e^\vartheta < 3$ (так как $\vartheta < 1$). Отсюда:

$$r_n < \frac{3}{(n+1)!}.$$

Чтобы вычислить значение e с точностью до 4-го десятичного знака, необходимо выбрать такое n, при котором $r_n < 0{,}0001$, т. е.

$$\frac{3}{(n+1)!} < \frac{1}{10000},\ \frac{(n+1)!}{3} > 10000,\ (n+1)! > 30000.$$

Это выполняется уже при $n = 7$. Имеем:

$$r_n < \frac{3}{8!} < 0{,}0001 .$$

Итак,

$$e \approx 1 + \frac{1}{1!} + \frac{1}{2!} + \frac{1}{3!} + \frac{1}{4!} + \frac{1}{5!} + \frac{1}{6!} + \frac{1}{7!} = 2{,}7181 .$$

Для получения большей точности следует взять большее n. Возможность достижения произвольной точности обеспечивается условием $\lim\limits_{n\to\infty} r_n = 0$. Принимая во внимание этот факт, можно записать:

$$e = 1 + \frac{1}{1!} + \frac{1}{2!} + \ldots + \frac{1}{n!} + \ldots$$

то есть число e является суммой бесконечного числового ряда

$$1 + \frac{1}{1!} + \frac{1}{2!} + \ldots + \frac{1}{n!} + \ldots = \sum_{n=0}^{\infty} \frac{1}{n!} .$$

Разложение функции e^x в степенной ряд.
Рассмотрим снова разложение

$$e^x = 1 + \frac{1}{1!} x + \frac{1}{2!} x^2 + \ldots + \frac{1}{n!} x^n + r_n \quad (x \neq 1)$$

$$r_n = \frac{e^{\vartheta x}}{(n+1)!} x^{n+1} = e^{\vartheta x} \frac{x^{n+1}}{(n+1)!} .$$

Поскольку $\vartheta < 1$, то для произвольного x

$$| e^{\vartheta x} | < e^x .$$

С другой стороны, как мы знаем, $\dfrac{x^{n+1}}{(n+1)!} \to 0$.

Следовательно, при любом x $\lim\limits_{n\to\infty} r_n = 0$.

Таким образом, мы показали, что функция e^x представима в виде суммы степенного ряда:

$$e^x = 1 + \frac{1}{1!} x + \frac{1}{2!} x^2 + \ldots + \frac{1}{n!} x^n + \ldots$$

Если, например, мы хотим вычислить с точностью до четвертого десятичного знака значение e^3, то n нужно выбрать еще большим, чем при вычислении e. Причина заключается в том, что, как нетрудно убедиться, при увеличении x скорость сходимости r_n к нулю уменьшается.

Разложение тригонометрических функций в степенные ряды. Аналогично можно получить следующие разложения:

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

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

Эти ряды сходятся при любом x. Ряд

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

сходится при $|x| < 1$.

Вообще бесконечно дифференцируемая функция $f(x)$ может быть разложена в бесконечный ряд Тейлора (степенной ряд) для всех тех x, для которых

$$\lim_{n\to\infty} r_n(x) = 0 .$$

Множество таких значений x — аналогично области сходимости степенного ряда — всегда является симметричным интервалом с центром в некоторой точке a.

Дифференцирование степенных рядов. Пусть дан равномерно сходящийся ряд функций. Если в результате его почленного дифференцирования мы вновь получим равномерно сходящийся ряд, то сумма полученного ряда будет производной суммы исходного ряда. Отсюда следует, что степенной ряд можно дифференцировать почленно. При этом интервал сходимости ряда производных совпадает с интервалом сходимости исходного ряда функций.

Пример:

$$\ln(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + \frac{x^5}{5} - \frac{x^6}{6} + \ldots \quad (|x| < 1)$$

В результате почленного дифференцирования получаем:

$$\frac{1}{1+x} = 1 - x + x^2 - x^3 + x^4 - x^5 + \ldots .$$

Этот же результат можно получить суммированием геометрического ряда, стоящего в правой части равенства.

НЕОПРЕДЕЛЕННЫЙ ИНТЕГРАЛ

$y = x^2$; $y' = 2x$. Можно сказать, что функции $y = x^2$ поставлена в соответствие другая функция: ее производная. Функция $2x$ является производной функции x^2. Наоборот, функцию x^2 называют первообразной функции $2x$. Вообще, если

$$F'(x) = f(x),$$

то функцию $F(x)$ называют *первообразной* по отношению к $f(x)$.

Первообразной от функции $2x$ является не только функция x^2, но и функции $x^2 + 1$, $x^2 + 2$, ... и вообще функции вида $x^2 + C$, где C — произвольная постоянная. То же самое верно и в общем случае: если $F(x)$ — первообразная от функции $f(x)$, то первообразными от $f(x)$ будут все функции вида $F(x) + C$, и только они.

Отыскание первообразной для данной функции называется неопределенным интегрированием и обозначается следующим образом:

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

Например:

$$\int 2x\,dx = x^2 + C.$$

Константа C называется *постоянной интегрирования.*

ТАБЛИЦА ОСНОВНЫХ ИНТЕГРАЛОВ

Всякая формула дифференцирования, рассмотренная в обратном порядке, дает нам формулу интегрирования, например:

$$(x^2 + \cos x + e^{2x})' = 2x - \sin x + 2e^{2x}$$

$$\int (2x - \sin x + 2e^{2x})dx = x^2 + \cos x + e^{2x} + C.$$

Таким образом из формул дифференцирования основных элементарных функций нетрудно получить так называемую таблицу (основных) простейших интегралов:

$$1.\ \int x^\alpha\,dx = \begin{cases} \dfrac{x^{\alpha+1}}{\alpha+1} + C, & \text{если}\quad \alpha \neq -1 \\ \ln x + C, & \qquad\ \ \alpha = -1 \end{cases}$$

$$2.\ \int e^x\,dx = e^x + C\,; \qquad \int a^x\,dx = \frac{a^x}{\ln a} + C$$

$$3.\ \int \sin x\ dx = -\cos x + C\,; \qquad \int \cos x\ dx = \sin x + C$$

$$\int \frac{dx}{\cos^2 x} = \operatorname{tg} x + C\,; \qquad \int \frac{dx}{\sin^2 x} = -\operatorname{ctg} x + C$$

$$4.\ \int \frac{dx}{\sqrt{1-x^2}} = \begin{cases} \arcsin x + C\,; \\ -\arccos x + C \end{cases} \int \frac{dx}{1+x^2} = \begin{cases} \operatorname{arc\,tg} x + C \\ -\operatorname{arc\,ctg} x + C \end{cases}$$

$$5.\ \int \operatorname{sh} x\,dx = \operatorname{ch} x + C\,; \qquad \int \operatorname{ch} x\,dx = \operatorname{sh} x + C$$

$$\int \frac{dx}{\operatorname{ch}^2 x} = \operatorname{th} x + C\,; \qquad \int \frac{dx}{\operatorname{sh}^2 x} = -\operatorname{cth} x + C.$$

Естественно, что эти формулы справедливы лишь на определенных множествах. Так, например, формула $\int \frac{dx}{\sqrt{1-x^2}} = \arcsin x + C$ действительна только в интервале $(-1\,;\ 1)$.

ПРАВИЛА НЕОПРЕДЕЛЕННОГО ИНТЕГРИРОВАНИЯ

Путем обращения правил дифференцирования можно получить так называемые правила интегрирования. Перечислим некоторые из них.

1. Сумму можно интегрировать почленно.
2. Постоянный множитель можно выносить из-под знака интеграла.

Пример:

$$\int (2x^2 - 3x + 5)dx = 2\int x^2 dx - 3\int x\ dx + 5\int dx = 2\frac{x^3}{3} - 3\frac{x^2}{2} +$$
$$+ 5x + C = \frac{2}{3}x^3 - \frac{3}{2}x^2 + 5x + C\,.$$

3. Рассмотрим правило дифференцирования произведения:

$$\boxed{(u \cdot v)' = u'v + u \cdot v'\,.}$$

Отсюда:

$$\boxed{uv = \int u'v\,dx + \int uv'dx,}$$

т. е.

$$\boxed{\int u'v\,dx = uv - \int v'u\,dx + C\,.}$$

Это так называемое *правило дифференцирования по частям*. Суть его в том, что вычисление интеграла

$$\int u'v\,dx$$

сводится к вычислению

$$\int v'u\,dx\,.$$

Этой формулой имеет смысл пользоваться в тех случаях, когда второй интеграл проще вычислять, чем первый. Отсюда следует, что не все равно, который из сомножителей следует выбрать в качестве u', а который — в качесте v. Заметим однако, что правило интегрирования по частям применимо не всегда.

Примеры:

1. $$\int xe^x\,dx$$

$$vu'$$

Пусть

$$\left.\begin{aligned} e^x &= u \\ x &= v \end{aligned}\right\}, \quad \text{тогда} \quad \left\}\begin{aligned} u' &= e^x \\ v' &= 1\,, \end{aligned}\right.$$

Отсюда:

$$\int xe^x dx = e^x x = \int 1 \cdot e^x dx = e^x x - e^x + C.$$

2. $$\int x \sin x\,dx$$

$$vu'.$$

Пусть

$$\left.\begin{aligned} u &= -\cos x \\ v &= x \end{aligned}\right\}, \quad \text{тогда} \quad \left\{\begin{aligned} u' &= \sin x \\ v' &= 1 \end{aligned}\right. .$$

Отсюда:

$$\int x \sin x\,dx = -x\cos x + \int 1 \cdot \cos x\,dx = -x\cos x + \sin x + C\,.$$

ИНТЕГРИРОВАНИЕ С ПОМОЩЬЮ ЗАМЕНЫ ПЕРЕМЕННОЙ

В результате обращения правила дифференцирования сложной функции мы получаем следующую формулу:

$$\int f(x)\,dx = \left[\int f[\varphi(t)]\varphi'(t)\,dt\right]_{t=\psi(x)}.$$

При этом в подинтегральном выражении производится замена переменной $x = \varphi(t)$, добавляется множитель $\varphi'(t)$ и совершается интегрирование по t, после чего следует произвести обратную замену переменной $t = \psi(x)$, где $\psi(x)$ — функция, обратная к $\varphi(t)$. Поскольку при обратной замене переменной используется функция, обратная к $\varphi(t)$, то $\varphi(t)$ должна быть монотонной.

Справедливость записанной выше формулы замены переменной под знаком интеграла легко может быть проверена дифференцированием. Формально замену можно произвести, написав вместо x $\varphi(t)$, а вместо дифференциала dx

$$dx = \varphi'(t)\, dt\,.$$

В этом случае мы перейдем к интегрированию по t.

Пример:

$$\int \sqrt{1-x^2}\, dx$$

Пусть

$$x = \cos t$$

$$dx = -\sin t\, dt$$

$$t = \arccos x$$

$$\int \sqrt{1-x^2}\, dx = \int \sqrt{1-\cos^2 t} \cdot (-\sin t) dt = -\int \sin^2 t\, dt =$$

$$= -\int \frac{1-\cos 2t}{2}\, dt = \frac{\sin 2t}{4} - \frac{t}{2} + C = \frac{x\sqrt{1-x^2} - \arccos x}{2} + C.$$

ИНТЕГРИРОВАНИЕ РАЦИОНАЛЬНЫХ ФУНКЦИЙ

Пусть требуется вычислить

$$\int \frac{dx}{x^2-1}\,.$$

Знаменатель подинтегральной функции можно представить в виде произведения:

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

Определим теперь постоянные A и B, удовлетворяющие следующему равенству:

$$\frac{1}{x^2-1} = \frac{A}{x-1} + \frac{B}{x+1}\,.$$

Требуется найти A и B, то есть две неизвестные величины. Для этого нужны два уравнения, в то время как нам, казалось бы, дано одно:

$$\frac{1}{x^2-1}=\frac{A}{x-1}+\frac{B}{x+1}.$$

Однако если принять во внимание, что наше уравнение справедливо при всех значениях x, то нетрудно получить два уравнения, подставив в данное два различных значения x. Пусть

$$\begin{array}{ll} x=0 & -1=-A+B \\ x=2 & \dfrac{1}{3}=A+\dfrac{B}{3} \\ \hline \end{array}$$

Получаем: $$A=\frac{1}{2},\ B=-\frac{1}{2}.$$

Отсюда:

$$\int\frac{dx}{x^2-1}=\frac{1}{2}\int\frac{dx}{x-1}-\frac{1}{2}\int\frac{dx}{x+1}=\frac{1}{2}\ln(x-1)-\frac{1}{2}\ln(x+1)+C.$$

Этот прием может быть использован при интегрировании любых дробных рациональных функций, знаменатели которых могут быть разложены на простые множители.

ЗАМЕЧАНИЕ О ДИФФЕРЕНЦИРОВАНИИ И ИНТЕГРИРОВАНИИ

Интегрирование, как правило, значительно сложнее дифференцирования. Оно не является настолько механическим, требует большей практики и изобретательности. Мы лишь кратко обрисовали методы интегрирования. Существуют многочисленные таблицы интегралов, содержащие систематические и подробные изложения этих методов.

Следует, однако, выделить одно существенное принципиальное различие между дифференцированием и интегрированием.

Производная всякой элементарной функции есть также элементарная функция, иногда, правда, сложная. В то же время нетрудно указать пример такой элементарной функции, интеграл от которой уже не является функцией элементарной. Таковы, например, функции

$$\frac{1}{\ln x},\quad \frac{a^x}{x},\quad \frac{\sin x}{x}.$$

Это означает, что нет такой элементарной функции $F(x)$, для которой

$$F'(x) = \frac{1}{\ln x}.$$

Такая функция, разумеется, существует, только уже не является элементарной. На самом деле можно доказать, что если функция $f(x)$ непрерывна в некотором интервале $[a;\ b]$, то всегда существует такая функция $F(x)$, что

$$F'(x) = f(x).$$

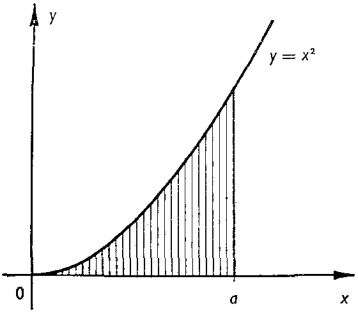
Только, например, для функции $f(x) = \dfrac{1}{\ln x}$ функция $F(x)$ оказывается не элементарной.

ОПРЕДЕЛЕННЫЙ ИНТЕГРАЛ

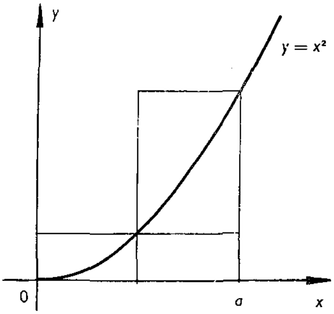
ОПРЕДЕЛЕНИЕ ИНТЕГРАЛА

Вычислим или, точнее, определим площадь плоской фигуры, лежащей между дугой параболы $y = x^2$ и отрезком оси $0x$ от $x = 0$ до $x = a$. При этом мы будем исходить из следующей аксиоматической предпосылки: площадь фигуры есть число, которое больше, чем площадь любой вписанной в нее элементарной фигуры, ограниченной отрезками прямых, и меньше, чем площадь любой описанной элементарной фигуры.

В качестве вписанной и описанной элементарных фигур рассмотрим фигуры, составленные из прямоугольников. Разобьем отрезок $[0;\ a]$ оси $0x$ на n равных частей. Полученные интервалы разбиения будут служить основаниями прямоугольников. Для вписанной фигуры высоту каждого прямоугольника будем выбирать равной минимальному значению функции $y = x^2$ в соответствующем интервале разбиения, а для описанной



Площадь фигуры под параболой $y = x^2$

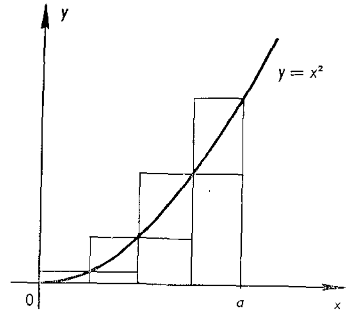


Вписанная и описанная элементарные фигуры в случае $n = 2$

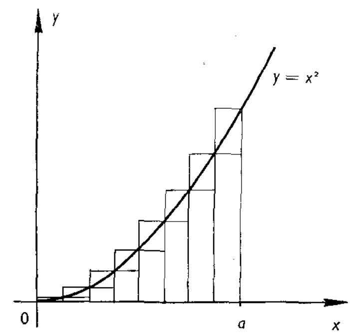
фигуры — равной максимальному значению. Площади вписанной и описанной элементарных фигур будут равны суммарным площадям соответствующих прямоугольников. Изобразим элементарные фигуры, получающиеся в случае $n = 2$.

Минимальное значение функции на отрезке $[0;\ a]$ равно 0. Значит, при построении мы можем получить прямоугольник с высотой 0. Его площадь, естественно, есть основание $\cdot\, 0 = 0$.

На рисунках изображены элементарные фигуры для $n = 4$ и $n = 8$.



Вписанная и описанная элементарные фигуры при $n = 4$



Вписанная и описанная элементарные фигуры при $n = 8$

Рисунки для случаев $n = 2, 4, 8$ хорошо иллюстрируют тот факт, что при неограниченном возрастании n (числа интервалов разбиения) площади вписанных и описанных элементарных фигур стремятся к одному и тому же пределу.

Принимая во внимание, что

$$1^2 + 2^2 + \ldots + n^2 = \frac{n(n+1)(2n+1)}{6},$$

получаем формулы для нижней (соответствующей вписанной фигуре) и верхней (соответствующей описанной фигуре) сумм площадей прямоугольников (их называют нижней и верхней суммами Дарбу):

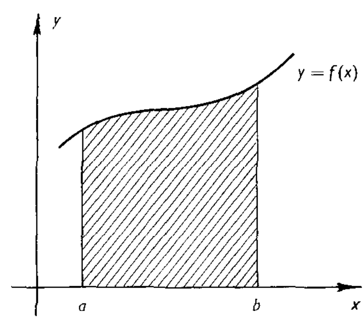
$$\underline{\Sigma}_n = a^3\,\frac{2n^3 + 3n^2 + n}{6n^3}$$

$$\overline{\Sigma}_n = a^3\,\frac{2n^3 - 3n^2 + n}{6n^3}.$$

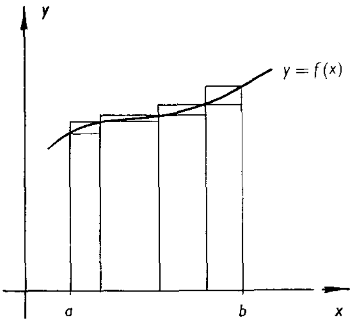
Отсюда:

$$\boxed{\lim_{n\to\infty} \underline{\Sigma}_n = \lim_{n\to\infty} \overline{\Sigma}_n = \frac{a^3}{3}.}$$

Тем самым доказано утверждение, высказанное нами из наглядных соображений. Искомая площадь равна $\frac{a^3}{3}$, так как это единственное число, которое больше, чем пло-



Криволинейная трапеция, ограниченная сверху графиком функции $y = f(x)$, а с боков — ординатами этого графика при $x = a$ и $x = b$



Разбиение отрезка $[a;\ b]$ оси x на n частей

щадь любой вписанной, и меньше, чем площадь любой описанной элементарной фигуры.

Рассмотрим теперь задачу в общем виде: требуется определить площадь криволинейной трапеции, то есть фигуры, ограниченной снизу отрезком $[a;\ b]$ оси $0x$, сверху — графиком некоторой функции

$$y = f(x)\,,$$

с боков — ординатами этого графика при $x = a$ и $x = b$.

Разобьем отрезок $[a;\ b]$ на n частей, длины которых равны, соответственно,

$$\Delta x_1,\ \Delta x_2,\ \ldots,\ \Delta x_n\,.$$

Обозначение Δx и здесь указывает на вычисление разности. Действительно, разбиение производится с помощью точек деления, к которым относятся и крайние точки a и b. Пусть точками деления будут

$$a = x_0,\ x_1,\ x_2,\ \ldots,\ x_{n-1},\ x_n = b\,.$$

Тогда

$$\Delta x_1 = x_1 - x_0,\ \Delta x_2 = x_2 - x_1,\ \ldots\ \Delta x_n = x_n - x_{n-1}\,.$$

Выше мы для каждого такого разбиения строили нижнюю и верхнюю суммы Дарбу:

$$\underline{\Sigma}_n = \sum_{i=1}^{n} \Delta x_i\, m_i,\quad \overline{\Sigma}_n = \sum_{i=1}^{n} \Delta x_i\, M_i\,,$$

где m_i и M_i обозначают, соответственно, минимальное и максимальное значения функции $f(x)$ на отрезке $[x_{i-1};\ x_i]$. Эти суммы и в общем случае должны находиться в такой же связи с площадью криволинейной трапеции, как в рассмотренном выше примере: площадь криволинейной трапеции должна быть не меньше любой нижней и не больше любой верхней суммы Дарбу. Если существует единственное число, удовлетворяющее этим условиям, то это число следует принять за площадь трапеции. Можно доказать (здесь мы не будем не этом останавливаться), что если $f(x)$ — непрерывная функция, то при неограниченном измельчении разбиения нижняя и верхняя суммы Дарбу сходятся к одному пределу. В этом случае предел определяет площадь соответствующей криволинейной трапеции.

Что означает неограниченное измельчение разбиения? Если отрезок $[a;\ b]$ разбить на n равных частей, как мы это сделали в примере, то смысл этого выражения ясен: $n \to \infty$. Однако в общем случае мы не требуем, чтобы отрезки $\Delta x_1, \Delta x_2, \ldots, \Delta x_n$ были равны. Важно, чтобы и наибольший из них становился произвольно малым.

Подведем итог: если задана такая последовательность разбиений, что длина наибольшего отрезка разбиения стремится к 0, то верхняя и нижняя суммы Дарбу сходятся к одному пределу. Этот предел называется определенным интегралом от функции $f(x)$ на отрезке $[a;\ b]$ и обозначается символом

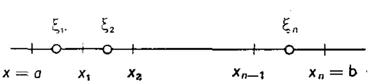
$$\int_a^b f(x)\,dx\,.$$

Следует заметить, что если функция $f(x)$ не является непрерывной, то нижяя и верхняя суммы Дарбу не всегда сходятся к одному пределу. В таком случае говорят, что функция $f(x)$ не интегрируема на отрезке $[a;\ b]$ (Символ $\int$ — поскольку речь идет о пределе суммы — происходит от первой буквы латинского слова summa (сумма)). a и b называются, соответственно, нижним и верхним пределами интеграла, а функция $f(x)$ — подинтегральной функцией. Во избежание ошибок следует подчеркнуть, что точно так же, как символ дифференцирования $\dfrac{dy}{dx}$ не означает деления, подинтегральное выражение

$$f(x)\ dx$$

не означает умножения.

Данное определение интеграла можно несколько обобщить. Дело в том, что при построении сумм не обязательно выбирать минимальные и максимальные значения функции в отдельных интервалах разбиения $\Delta x_1, \Delta x_2, \ldots, \Delta x_n$, достаточно в каждом интервале выбрать произвольную точку: в первом — ξ_1, во втором — ξ_2 и так далее. Тогда в качестве высот прямоугольников с основаниями $\Delta x_1, \Delta x_2, \ldots, \Delta x_n$ следует, соответственно взять $f(\xi_1), f(\xi_2), \ldots, f(\xi_n)$. Поскольку



В каждом отрезке разбиения интервала $[a;\ b]$ оси x выберем произвольную точку ξ;

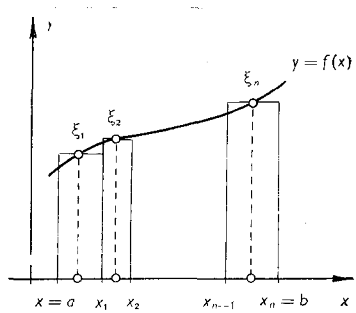
$$m_i \leqq f(\xi) \leqq M_i\,,$$

то очевидно, что величина получаемой таким образом интегральной суммы

$$\Sigma_n = \sum_{i=1}^{n} \Delta x_i\, f(\xi_i)$$

лежит между величинами сумм Дарбу, соответствующих тому же разбиению:

$$\sum_{i=1}^{n} \Delta x_i\, m_i = \underline{\Sigma}_n \leqq \Sigma_n =$$
$$= \sum_{i=1}^{n} \Delta x_i\, f(\xi_i) \leqq \overline{\Sigma}_n = \sum_{i=1}^{n} \Delta x_i\, M_i\,.$$



составим интегральную сумму

Отсюда следует, что при неограниченном измельчении разбиения предел интегральной суммы

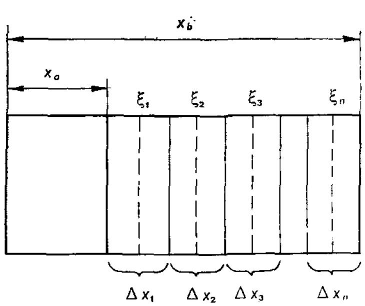
$$\sum_{i=1}^{n} \Delta x_i\, f(\xi_i)$$

должен совпасть с пределом нижней и верхней сумм Дарбу, т. е.

$$\boxed{\int_a^b f(x)\,dx = \lim_{\Delta x_i \to 0} \sum_{i=1}^{n} \Delta x_i\, f(\xi_i)\,.}$$

При введении понятия определенного интеграла мы воспользовались его геометрической трактовкой. Тем не менее, как хорошо видно из последней формулы, это понятие не геометрическое. Определенный интеграл — это предел специального типа.

Рассмотрим пример из области физики. Пусть требуется определить величину работы, которую совершает при расширении сжатый газ, находящийся в цилиндре с поршнем. Мерой расширения газа может служить величина сдвига поршня. Для простоты предположим, что площадь поверхности поршня равна 1. Тогда, если p — давление газа, то и величина силы, с которой газ давит на поршень, равна p. Обозначим расстояние от поршня до края цилиндра через x. Величину L совершенной работы мы получим, если величину силы — в данном случае p — умножим на величину $x_b - x_a$ сдвига поршня, то есть $L = p(x_b - x_a)$. Однако эта формула оказывается приближенной, так как пока поршень из x_a передвинется в x_b, величина давления p изменится. Действительно, различным положениям x поршня отвечают разные величины давления, то есть давление является функцией x: $p = p(x)$. Здесь опять может помочь уже хорошо знакомый нам метод. Действительно, если поршень передвинется на очень малое расстояние Δx, то давление при этом изменится незначительно. Значит, если ξ_1 — произвольная точка отрезка Δx_1, то приближенно можно считать, что давление на всем отрезке Δx_1 равно $p(\xi_1)$. Следовательно, работа, совершаемая газом по передвижению поршня на расстояние Δx_1, приближенно равна $p(\xi_1)\Delta x_1$. Из конца отрезка Δx_1 отложим отрезок Δx_2, на котором изменение давления газа также незначительно, то есть если ξ_2 — произвольная точка отрезка Δx_2, то давление на всем отрезке x_2 можно считать равным $p(\xi_2)$. Тогда работа газа по передвижению поршня из конца отрезка Δx_1 на расстояние Δx_2 прибли-



Работа, которую совершает при расширении сжатый газ, находящийся в цилиндре с поршнем женно равна $p(\xi_2) \cdot \Delta x_2$. Продолжая этот процесс дальше, можно путем малых передвижений $\Delta x_1, \Delta x_2, \ldots, \Delta x_n$ перейти из x_a в x_b. Работа, совершаемая газом на этом пути, приближенно равна

$$\text{(I)} \qquad L \approx p(\xi_1)\Delta x_1 + p(\xi_2)\Delta x_2 + \ldots + p(\xi_n)\Delta x_n .$$

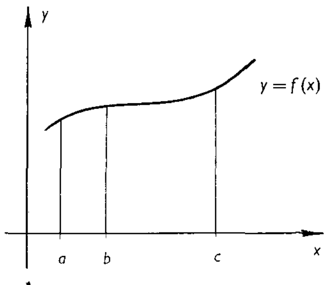
Очевидно, что эта величина будет тем точнее, чем мельче разбиение отрезка $x_b - x_a$. Следовательно, точное значение величины совершенной работы можно получить, перейдя в (I) к пределу при неограниченном измельчении разбиения отрезка $[x_a;\ x_b]$, которое означает, что длина наибольшего из отрезков разбиения неограниченно уменьшается. Предел выражения в правой части (I), как мы уже видели, обозначается символом

$$\int_{x_a}^{x_b} p(x)\,dx,$$

т. е.

$$L = \lim_{\Delta x_i \to 0} \sum_{i=1}^{n} p(\xi_i)\,\Delta x_i = \int_{x_a}^{x_b} p(x)\,dx.$$

Предел суммы, стоящей в правой части (I), который дает нам таким образом величину работы, совершаемой расширяющимся газом, мы назвали определенным интегралом от функции $p(x)$ на отрезке $[x_a;\ x_b]$.

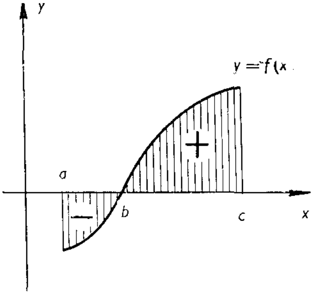


Аддитивность определенного интеграла

Некоторые важные свойства определенного интеграла. Таким образом мы пришли к понятию определенного интеграла другим путем. Отметим, что существует еще много теоретических и практических задач, приводящих к пределам подобного типа. Однако введенное понятие еще не дает общего метода вычисления определенных интегралов, а следовательно, и способа решения задач, приводящих к ним (как, например, вычисление площадей криволинейных трапеций). Тем не менее, мы уже можем установить несколько важных свойств определенных интегралов.

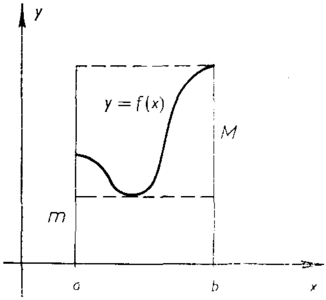
1. *Аддитивность.*

$$\int_a^b f(x)\,dx + \int_b^c f(x)\,dx = \int_a^c f(x)\,dx.$$



Определение интеграла без изменений переносится на случай функции $f(x)$, имеющей произвольный знак

2. *Интеграл от функции, имеющей произвольный знак.* До сих пор для наглядности интегрируемая функция $f(x)$ на всех графиках была положительной. Однако при построении интегральных сумм это не играет роли.



Значение интеграла от функции $y = f(x)$ заключено между $m(b-a)$ и $M(b-a)$, $\int f(x)\,dx = f(\xi)(b-a)$

Определение интеграла без изменений распространяется на функцию $f(x)$, имеющую произвольный знак. В этом случае интеграл дает площадь криволинейной трапеции от a до b с минусом, а от b до c — с плюсом. В результате мы получаем разность этих площадей.

3. *Перемена пределов интеграла.* Целесообразно определить интеграл

$$\int_a^b f(x)\,dx$$

и в случае, когда $a \geqq b$. Получаем:

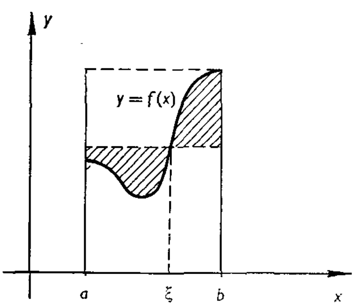
$$\int_a^b f(x)\,dx = -\int_b^a f(x)\,dx.$$

Если $a = b$, то

$$\int_a^a f(x)\,dx = 0.$$

4. *Теорема о среднем значении.* Значение интеграла заключено между $m(b-a)$ и $M(b-a)$, где m и M обозначают, соответственно, минимальное и максимальное значения функции $f(x)$ на отрезке $[a;\ b]$. (Все интегральные суммы заключены в этих границах, а следовательно и интеграл):

$$m(b-a) \leqq \int_a^b f(x)\,dx \leqq M(b-a).$$



Интегральная теорема о среднем значении

Если функция $f(x)$ непрерывна, то внутри интервала $[a;\ b]$ существует такая точка ξ, в которой функция принимает значение $\frac{1}{b-a}\int_b^a f(x)\,dx$. Следовательно,

$$\boxed{\int_a^b f(x)\,dx = f(\xi)(b-a).}$$

Это интегральная теорема о среднем значении.

СВЯЗЬ МЕЖДУ ДИФФЕРЕНЦИАЛЬНЫМ И ИНТЕГРАЛЬНЫМ ИСЧИСЛЕНИЯМИ

ОСНОВНАЯ ТЕОРЕМА ДИФФЕРЕНЦИАЛЬНОГО И ИНТЕГРАЛЬНОГО ИСЧИСЛЕНИЯ

Как мы видели из предыдущего, и в самой математике, и в ее приложениях чрезвычайно важную роль играют два типа пределов. Один из них:

$$\lim_{h\to 0}\frac{f(x+h)-f(x)}{h}=f'(x)$$

дает нам производную, другой —

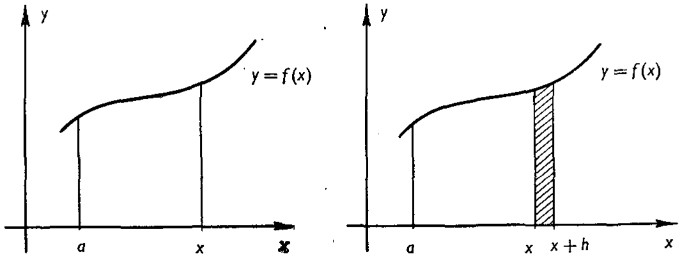
$$\lim_{\Delta x\to 0}\sum_{i=1}^{n}\Delta x_i\, f(\xi_i)=\int_a^b f(x)\,dx$$

— определенный интеграл от функции. Очень важное значение имеет теорема, устанавливающая связь между двумя этими понятиями. Для того, чтобы перейти к ней, рассмотрим определенный интеграл как функцию его верхнего предела:

$$F(x)=\int_a^x f(t)\,dt$$

(в этом случае переменную интегрирования следует обозначать другой буквой). Основная теорема утверждает следующее:

если $f(x)$ — непрерывная функция, то $F'(x)=f(x)$, т. е. рассматривая интеграл от непрерывной функции как функцию верхнего предела, мы тем самым получаем некоторую первообразную от интегрируемой функции.



Основная теорема дифференциального и интегрального исчисления

ВЫЧИСЛЕНИЕ ОПРЕДЕЛЕННОГО ИНТЕГРАЛА С ПОМОЩЬЮ НЕОПРЕДЕЛЕННОГО ИНТЕГРИРОВАНИЯ

Пусть

$$F(x) = \int_a^x f(t)\,dt,$$

т. е.

$$F'(x) = f(x)$$

и пусть $G(x)$ — некоторая первообразная от функции $f(x)$, то есть

$$G'(x) = f(x).$$

Тогда функции $F(x)$ и $G(x)$, будучи первообразными от одной и той же функции $f(x)$, различаются лишь на аддитивную постоянную:

$$F(x) = G(x) + C.$$

Подставляя в это равенство $x = a$, получаем:

$$C = -G(a), \qquad F(x) = G(x) - G(a)$$

$$\int_a^x f(t)\,dt = G(x) - G(a),$$

$$\int_a^b f(t)\,dt = G(b) - G(a).$$

Последнее равенство часто записывают в такой форме:

$$\boxed{\int_a^b f(x)\,dx = [G(x)]_a^b.}$$

Таким образом, определенный интеграл может быть вычислен с помощью одной первообразной. Если известна некоторая функция $G(x)$, являющаяся первообразной по отношению к функции $f(x)$, то определенный интеграл $\int_a^b f(x)\,dx$ можно вычислить следующим образом: в функцию $G(x)$ следует подставить сначала верхний предел (b), затем нижний предел (a) интеграла и вычислить разность $G(b) - G(a)$.

Применим это правило для решения нескольких примеров.

1. $$\int_1^3 x^3\,dx = \left[\frac{x^4}{4}\right]_1^3 = \frac{81}{4} - \frac{1}{4} = \frac{80}{4} = 20.$$

2. $$\int_1^2 \frac{1}{x} dx = [\ln x]_1^2 = \ln 2 - \ln 1 = \ln 2.$$

3. $$\int_0^\pi \sin x \, dx = [-\cos x]_0^\pi = 1 - (-1) = 2.$$

ПРИМЕНЕНИЕ ИНТЕГРАЛЬНОГО ИСЧИСЛЕНИЯ ДЛЯ НАХОЖДЕНИЯ ДЛИНЫ ДУГИ

Длину дуги кривой $y = f(x)$ от $x = a$ до $x = b$ мы определим как предел длин вписанных в нее ломаных линий при неограниченном измельчении разбиения. Таким образом мы придем к следующей формуле:

$$\boxed{l = \int_a^b \sqrt{1 + [f'(x)]^2}\, dx.}$$

Пример:

Вычислить длину полуокружности радиуса 1.

$$f(x) = \sqrt{1 - x^2}, \qquad f'(x) = -\frac{1}{\sqrt{1 - x^2}}$$

$$1 + [f'(x)]^2 = \frac{1}{1 - x^2}.$$

Длина полуокружности есть

$$\int_{-1}^{1} \frac{dx}{\sqrt{1 - x^2}} = [\arcsin x]_{-1}^{1} = \arcsin 1 - \arcsin(-1) = \pi.$$

Следовательно, длина окружности радиуса 1 равна 2π. Отсюда получаем, что длина окружности радиуса r равна $2r\pi$.

ФУНКЦИИ НЕСКОЛЬКИХ ПЕРЕМЕННЫХ

Область определения функции. Функция $y = f(x)$ ставит в соответствие определенным вещественным числам x (всем числам, для которых $f(x)$ определена, множество H таких чисел называется областью определения функции) вещественные числа y, множество H' которых называют областью значений функции. Другими словами, функция отображает некоторое множество H вещественных чисел в некоторое другое множество вещественных чисел H'. Принимая во внимание известное соответствие между вещественными числами и точками числовой оси, можно сказать, что функция $y = f(x)$ отображает множество H точек некоторой прямой в множество H' точек некоторой другой прямой, то есть каждой точке $x \in H$ ставит в соответствие некоторую точку $y \in H'$. Нетрудно распространить понятие функции на случаи, когда область определения H и область значений H' являются множествами точек прямой, плоскости или трехмерного пространства.

Функции двух и трех переменных. Из аналитической геометрии известно, что точки плоскости можно задавать двумя вещественными числами, а точки трехмерного пространства — тремя, а именно, их координатами. Отсюда, например, следует, что если H — множество точек трехмерного пространства, а H' — множество точек некоторой прямой, то отображение в этом случае означает, что каждой тройке вещественных чисел, являющейся элементом множества H, ставится в соответствие некоторое число из множества H'. В этом случае говорят о функции трех переменных, которая обозначается так: $u = f(x; y; z)$. Если H — множество точек на плоскости, а H' — на прямой, то имеем функцию двух переменных: $z = f(x, y)$. Пусть H — множество точек на плоскости, а H' — множество точек на некоторой другой плоскости. Введем на плоскости множества H систему координат $(x; y)$, а на плоскости множества H' — систему $(u; v)$. В этом случае вместо обозначения функции $P' = f(P)$, где P обозначает точку с координатами $(x; y)$, а P' — точку $(u; v)$ часто употребляют следующую запись:

$$u = g(x, y) \qquad v = h(x, y).$$

Тем самым находит свое выражение тот факт, что рассматриваемая функция по существу эквивалентна двум вещественным функциям двух переменных. Аналогично, если H и H' — множества точек трехмерного пространства, то отображение $P' = f(P)$ множества H во множество H' можно задать тремя вещественными функциями трех переменных:

$$u = g(x; y; z); \qquad v = h(x; y; z); \qquad w = l(x; y; z).$$

В заключение отметим, что, поскольку три координаты определяют вектор в трехмерном пространстве, то как $(x; y; z)$, так и $(u; v; w)$ можно рассматривать как координаты некоторых векторов. В этом случае можно считать, что функция ставит в соответствие каждому вектору множества H (векторов) некоторый вектор из множества H'. В таком смысле говорят о векторных функциях векторного переменного. Употребляя эту терминологию, вещественную функцию трех переменных можно называть скалярной функцией векторного переменного.

ТОЧЕЧНЫЕ МНОЖЕСТВА

Понятия и теоремы, с которыми мы познакомились при изучении множеств, элементы которых — вещественные числа (точки прямой), могут быть путем ряда очевидных обобщений перенесены и на случай, когда элементами множеств являются точки плоскости или трехмерного пространства. Основным понятием при этом является расстояние между двумя точками, известное из аналитической геометрии:

$$d(P_1,P_2)=\sqrt{(x_1-x_2)^2+(y_1-y_2)^2} \quad \text{или}$$
$$d(P_1,P_2)=\sqrt{(x_1-x_2)^2+(y_1-y_2)^2+(z_1-z_2)^2}.$$

Очевидно, что оно удовлетворяет аксиомам расстояния:

1. $d(P_1, P_2)=0$ тогда и только тогда, когда $P_1=P_2$.
2. $d(P_1, P_2)=d(P_2, P_1)$ — свойство симметрии расстояния.
3. $d(P_1, P_3)\leqq d(P_1, P_2)+d(P_2, P_3)$ — неравенство треугольника.

Различные типы точечных множеств.

1. Множество точек называется ограниченным, если расстояние от его точек до начала координат ограничено (это означает, что множество целиком помещается в круге или шаре достаточно большого радиуса).

ε-окрестностью точки P называется множество всех точек x, которые удалены от P меньше, чем на ε, то есть для которых

$$d(Px)<\varepsilon.$$

ε-окрестностью точки P на плоскости является открытый круг радиуса ε, в пространстве — открытый шар того же радиуса с центром в точке P.

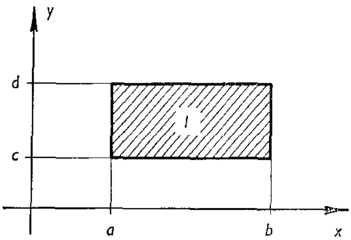
2. Множество называется открытым, если оно содержит всякую свою точку вместе с некоторой ее окрестностью.

Точка P называется предельной точкой множества H, если любая ее окрестность содержит бесконечно много точек из H.

3. Множество называется замкнутым, если оно содержит все свои предельные точки.

Прежде чем продолжить перечисление понятий, связанных с теорией точечных множеств, заметим, что очень часто рассматриваются окрестности другого типа.

Интервал. Интервалом на плоскости называется множество I точек, координаты $(x;\ y)$ которых удовлетворяют следующим неравенствам: $a\leqq x\leqq b,\quad c\leqq y\leqq d$. Если в каждом из записанных неравенств присутствует знак равенства, то интервал I — замкнутый, если же знак



Замкнутый интервал I на плоскости

равенства исключить, то мы получим открытый интервал. Совершенно аналогично неравенства

$$a \leq x \leq b, \qquad c \leq y \leq d, \qquad e \leq z \leq f$$

задают интервал в трехмерном пространстве.

Очевидно, что интервал на плоскости — это прямоугольник, а интервал в пространстве — параллелепипед.

Окрестностью точки P называют любой открытый интервал с центром в точке P. Можно доказать, что если вместо «сферических» ε-окрестностей рассмотреть «прямоугольные» окрестности, то для произвольного множества мы получим те же самые предельные точки, а следовательно, открытые множества останутся открытыми, замкнутые — замкнутыми. Этот факт позволяет назвать системы «сферических» и «прямоугольных» окрестностей эквивалентными.

Для множеств на плоскости и в пространстве остаются справедливыми теорема Больцано—Вейерштрасса и простейшие теоремы об открытых и замкнутых множествах, с которыми мы познакомились выше. Без изменений переносится на случай пространств, размерности которых больше единицы, и понятие плотного множества.

Множество называется *связным*, если не может быть разбито на два таких непересекающихся подмножества, ни одно из которых не содержит предельных точек другого. Открытое связное множество мы будем называть *областью*.

Некоторые важные понятия. Введем теперь несколько понятий, которые не упоминались при рассмотрении вещественных числовых множеств.

Точка $a \in H$ называется *внутренней точкой* множества H, если существует окрестность точки a, целиком лежащая в H. Следовательно, открытое множество характеризуется тем, что все его точки — внутренние.

Точка a называется *граничной точкой* множества H, если всякая ее окрестность содержит как точки из H, так и точки, не принадлежащие H. Так, например, множество замкнуто, если оно содержит все свои граничные точки, и открыто — если не содержит ни одной. Рассматривая некоторую область вместе с ее граничными точками, мы получим замкнутую область. Примером замкнутой области может служить замкнутый интервал.

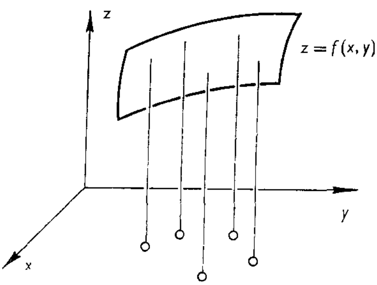
Теоремы и понятия, связанные с последовательностями — поскольку они основаны на понятии расстояния — также могут быть перенесены на случай пространств нескольких измерений. Заметим, что, например, в трехмерном пространстве сходимость последовательности $\{P_n\}$ (где P_n — общий член последовательности) к точке P_0 равносильна сходимости последовательностей координат $\{x_n\}$, $\{y_n\}$, $\{z_n\}$ к соотвествующим координатам x_0, y_0, z_0 точки P_0.

О ФУНКЦИЯХ НЕСКОЛЬКИХ ПЕРЕМЕННЫХ

Понятия предела и непрерывности функции остаются такими же, как и в случае функций одной вещественной переменной. Например, $\lim\limits_{P \to P_0} f(P) = A$ означает, что для любого числа $\varepsilon > 0$ найдется такое число $\delta > 0$, что если P находится внутри δ-окрестности точки P_0, то $f(P)$ попадает в ε-окрестность числа A. Функция $f(P)$ непрерывна в точке P_0, если существует предел функции в этой точке и этот предел равен $f(P_0)$.

ГРАФИК ФУНКЦИИ НЕСКОЛЬКИХ ПЕРЕМЕННЫХ

Пусть H — некоторое множество на плоскости, а $z = f(x, y)$ — вещественная функция, определенная на множестве H. Попробуем изобразить эту функцию графически. Для этого из точки $(x;\ y) \in H$ перпендикулярно плоскости xy (то есть параллельно оси z) построим отрезок длины $f(x, y)$ или, точнее, отметим в пространстве точку с координатами $(x;\ y;\ z = f(x, y))$. Если множество H является областью, а функция $f(x, y)$ непрерывна, то совокупность точек $(x;\ y;\ z = (fx, y))$ образует некоторую *поверхность*, которая задается уравнением $z = f(x, y)$.

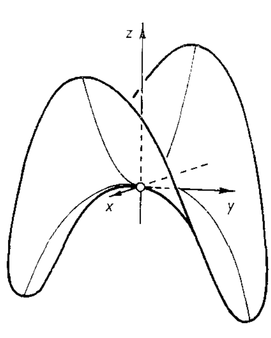


Поверхность $(x;\ y;\ z = f(x, y))$. Ее уравнение: $z = f(x, y)$

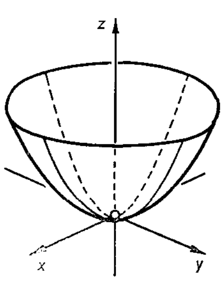
Примеры поверхностей:

1. $z = x^2 + y^2$; 2. $z = x^2 - y^2$.

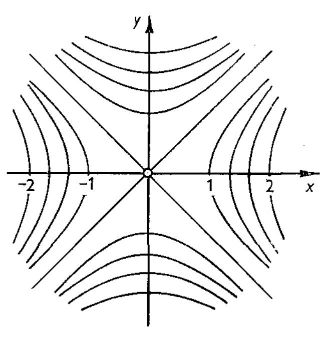
Рассмотрим поверхность, задаваемую некоторой непрерывной функцией $z = f(x, y)$. Если пересекать поверхность плоскостями, перпендикулярными оси z, то в сечении будут получаться некоторые кривые. Проектируя эти кривые на плоскость xy, мы будем получать так



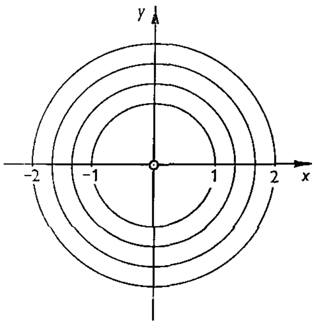
Поверхность $z = x^2 - y^2$



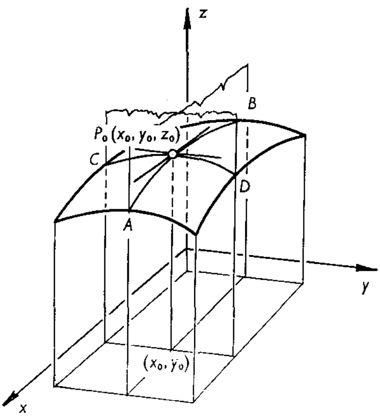
Поверхность $z = x^2 + y^2$



Линии уровней поверхности $z = x^2 - y^2$



Линии уровней поверхности $z = x^2 + y^2$



Плоскость $y = x_0$, перпендикулярная оси y, пересекает поверхность по кривой $z = f(x, y_0)$

называемые *линии уровней*, по виду которых можно судить о форме и «крутизне» поверхности. На каждой линии уровня функция $z = f(x, y)$ постоянна. (Поскольку сечение всякий раз производилось плоскостью, перпендикулярной оси z, на некоторой высоте $z = c$.) Если, например, последовательно пересекать поверхность плоскостями $z = c_0$, $z = c_1$, $z = c_2$, . . ., где числа c_n образуют некоторую арифметическую прогрессию (например, $c_0 = 0$, $c_1 = 1$, $c_2 = 2$, $c_3 = 3, \ldots$), то плотность получаемых линий уровня $f(x, y) = c_n$ $(n = 0, 1, 2, \ldots)$ позволяет сделать заключение о крутизне поверхности. Этот способ применяется при составлении географических карт для обозначения крутизны возвышенностей.

График функции, естественно, может быть определен и в том случае, если его не удается изобразить в виде «поверхности». Вообще под графиком вещественной функции $z = f(x, y)$, определенной на множестве H, понимается геометрическое место точек $(x; y; z)$ в пространстве, для которых $(x, y) \in H$ и $z = f(x, y)$. Хотя эта геометрическая интерпретация не применима в случае функций трех и большего числа переменных, тем не менее она оказывается весьма полезной, так как функции двух переменных обладают всеми теми свойствами и особенностями, которые отличают функции нескольких переменных от функций одной переменной.

Частные производные первого порядка. Если поверхность, задаваемую уравнением $z = f(x, y)$, пересечь плоскостью $y = y_0$, перпендикулярной к оси y и проходящей через некоторую точку $(x_0; y_0; z_0)$, то в сечении мы получим кривую $z = f(x, y_0)$ (z зависит только от x, y_0 — постоянная).

Производная функции $z = f(x, y_0)$ по переменной x называется *частной производной функции* $z = f(x, y)$ *по* x и обозначается символом $f'_x = \dfrac{\partial f}{\partial x}$.

Геометрически эта частная производная выражает угловой коэффициент касательной к кривой $z = f(x, y_0)$ в плоскости $y = y_0$. Аналогично определяется и частная производная по переменной y. Геометрически производная $f'_y = \dfrac{\partial f}{\partial y}$ выражает угловой коэффициент касательной к кривой $z = f(x_0, y)$.

Примеры:

1.
$$z = x^2y + \sin xy$$

$$f'_x = \frac{\partial z}{\partial x} = 2xy + y\cos xy \qquad f'_y = \frac{\partial z}{\partial y} = x^2 + x\cos xy$$

2.
$$z = x \cdot e^y$$

$$f'_x = \frac{\partial z}{\partial x} = e^y \qquad f'_y = \frac{\partial z}{\partial y} = xe^y.$$

Частные производные второго порядка. Определенные выше частные производные, естественно, сами являются функциями переменных x и y, а следовательно, также могут иметь частные производные. Таким образом, для функции $z = f(x, y)$ можно получить частные производные второго порядка:

$$f''_{xx} = \frac{\partial^2 f}{\partial x^2} = \frac{\partial}{\partial x}\left(\frac{\partial f}{\partial x}\right), \qquad \frac{\partial}{\partial y}\left(\frac{\partial f}{\partial x}\right) = \frac{\partial^2 f}{\partial x \partial y} = f''_{xy},$$

$$\frac{\partial}{\partial x}\left(\frac{\partial f}{\partial y}\right) = f''_{yx}, \qquad \frac{\partial}{\partial y}\left(\frac{\partial f}{\partial y}\right) = f''_{yy}.$$

Если смешанные производные f''_{xy} и f''_{yx} непрерывны, то они равны между собой, а значит, порядок дифференцирования не имеет значения.

Аналогичное утверждение оказывается верным и для смешанных производных более высокого порядка, а также для функций трех и большего числа переменных.

ПРОИЗВОДНЫЕ ОТ СЛОЖНЫХ ФУНКЦИЙ И ПОЛНЫЙ ДИФФЕРЕНЦИАЛ

Пусть функция $z = f(x, y)$ имеет непрерывные частные производные и пусть переменные x и y, в свою очередь, являются дифференцируемыми функциями некоторой переменной t (так называемого параметра). Тогда функция двух переменных $z = f(x, y)$ по существу является сложной функцией параметра t:

$$z = f[x(t); \quad y(t)].$$

При сделанных выше предположениях производная этой функции может быть вычислена по следующей формуле:

$$\boxed{\frac{dz}{dt} = \frac{\partial f}{\partial x} \cdot \frac{dx}{dt} + \frac{\partial f}{\partial y} \cdot \frac{dy}{dt}.}$$

Совершенно аналогично вычисляется производная сложной функции трех переменных $u = f(x, y, z)$:

$$\boxed{\frac{du}{dt} = \frac{\partial f}{\partial x} \cdot \frac{dx}{dt} + \frac{\partial f}{\partial y} \cdot \frac{dy}{dt} + \frac{\partial f}{\partial z} \cdot \frac{dz}{dt},}$$

где переменные x, y, z сами являются функциями параметра t.

Образуем дифференциал функции

$$z = z(t) = f[x(t);\ y(t)].$$

Как мы видели, дифференциал функции одной переменной равен произведению ее производной на дифференциал независимой переменной, то есть в данном случае

$$dz = \frac{dz}{dt}\, dt.$$

Следовательно,

$$\boxed{dz = \left(\frac{\partial f}{\partial x}\,\frac{dx}{dt} + \frac{\partial f}{\partial y}\,\frac{dy}{dt}\right) dt = \frac{\partial f}{\partial x}\,\frac{dx}{dt}\, dt + \frac{\partial f}{\partial y}\,\frac{dy}{dt}\, dt.}$$

Согласно определению дифференциала,

$$\boxed{\frac{dx}{dt}\, dt = dx, \qquad \frac{dy}{dt}\, dt = dy.}$$

Окончательно получаем:

$$\boxed{dz = \frac{\partial f}{\partial x}\, dx + \frac{\partial f}{\partial y}\, dy.}$$

Совершенно аналогично для функции трех переменных $u(x, y, z)$ имеем:

$$\boxed{du = \frac{\partial u}{\partial x}\, dx + \frac{\partial u}{\partial y}\, dy + \frac{\partial u}{\partial z}\, dz.}$$

Последние два дифференциала называются *полными дифференциалами* соответствующих функций нескольких переменных.

Если все производные в записанной формуле непрерывны, то полный дифференциал оказывается хорошим приближением для приращения функции.

Пример: Зависимость между сопротивлением R, напряжением E и силой тока I в электрической цепи задается равенством

$$R = \frac{E}{I}.$$

В результате произведенных измерений для E было получено значение 110 в с ошибкой измерения $dE = \pm 2$ в. Аналогично для I получилось значение 20 а с ошибкой $dI = \pm 0{,}5$ а. Пользуясь записанным выше соотношением, мы можем вычислить значение сопротивления со следующей ошибкой:

$$dR = \frac{\partial R}{\partial E} dE + \frac{\partial R}{\partial I} dI = \frac{1}{I} dE - \frac{E}{I^2} dI.$$

Чтобы оценить величину ошибки сверху, подставим в полученную формулу вместо результатов измерений их абсолютные величины. Таким образом мы получим верхнюю границу ошибки, совершаемой при измерении сопротивления:

$$|dR| \leq \frac{1}{20} \cdot 2 + \frac{110}{400} \cdot 0{,}5 \approx 0{,}24 \text{ ом}.$$

Полный дифференциал функции двух переменных. В рассматриваемых выше формулах для полного дифференциала частные производные сами являются функциями двух или трех переменных, то есть, например, полный дифференциал функции двух переменных имеет следующий вид:

$$P(x, y)\, dx + Q(x, y)\, dy.$$

В связи с этим встает вопрос, при каких условиях записанное линейное дифференциальное выражение является полным дифференциалом некоторой функции $z = f(x, y)$ двух переменных.

Предположим, что это дифференциальное выражение есть полный дифференциал некоторой функции $z = (x, y)$. В этом случае

$$\frac{\partial f}{\partial x} = P(x, y),$$

$$\frac{\partial f}{\partial y} = Q(x, y).$$

Продифференцировав первое из записанных равенств по y, а второе — по x, получаем:

$$\frac{\partial^2 f}{\partial y\,\partial x}=\frac{\partial P}{\partial y},$$

$$\frac{\partial^2 f}{\partial x\,\partial y}=\frac{\partial Q}{\partial x}.$$

Выше мы отмечали, что если смешанные производные непрерывны, то

$$\frac{\partial^2 f}{\partial y\,\partial x}=\frac{\partial^2 f}{\partial x\,\partial y},$$

поэтому

$$\frac{\partial P}{\partial y}=\frac{\partial Q}{\partial x}.$$

Это условие является необходимым для того, чтобы дифференциальное выражение $P\,dx+Q\,dy$ было полным дифференциалом некоторой функции $f(x,\ y)$ двух переменных. Оказывается, что при довольно общих предположениях оно же является и достаточным условием полного дифференциала. Итак, если в дифференциальном выражении $P\,dx+Q\,dy$ функции ($P(x,\ y)$ и $Q(x,\ y)$ непрерывны, и существуют их непрерывные частные производные по обеим переменным, то при определенных условиях, наложенных на соответствующую область плоскости xy и при выполнении в ней равенства $P'y=P'x$, существует такая функция двух переменных $f(x,\ y)$, для которой выражение $P\,dx+Q\,dy$ является полным дифференциалом.

Полный дифференция функции трех переменных. В случае функций трех переменных рассмотренная выше задача оказывается несколько сложней. Здесь также уместен вопрос, при каких условиях выражение

$$P(x,\ y,\ z)\,dx+Q(x,\ y,\ z)\,dy+R(x,\ y,\ z)\,dz$$

является полным дифференциалом некоторой функции трех переменных $u(x, y, z)$. Необходимое условие в этом случае задается тремя уравнениями. Действительно, если предположить, что выражение $P\,dx+Q\,dy+R\,dz$ есть полный дифференциал некоторой функции $u(x,\ y,\ z)$, то тогда

$$P(x,y,z)=\frac{\partial u}{\partial x};\qquad Q(x,y,z)=\frac{\partial u}{\partial y};\qquad R(x\,y,\ z)=\frac{\partial u}{\partial z}.$$

Продифференцировав первое уравнение отдельно по y и по z, второе — по x и по z, третье — по x и по y, получим, соответственно,

$$\frac{\partial P}{\partial y}=\frac{\partial^2 u}{\partial y\,\partial x},\qquad \frac{\partial P}{\partial z}=\frac{\partial^2 u}{\partial z\,\partial x};$$

$$\frac{\partial Q}{\partial x}=\frac{\partial^2 u}{\partial x\,\partial y}, \qquad \frac{\partial Q}{\partial z}=\frac{\partial^2 u}{\partial z\,\partial y};$$

$$\frac{\partial R}{\partial x}=\frac{\partial^2 u}{\partial x\,\partial z}, \qquad \frac{\partial R}{\partial y}=\frac{\partial^2 u}{\partial y\,\partial z}.$$

Поскольку смешанные производные равны (так как по предположению непрерывны), то имеем:

$$\frac{\partial P}{\partial y}=\frac{\partial Q}{\partial x}, \qquad \frac{\partial P}{\partial z}=\frac{\partial R}{\partial x}, \qquad \frac{\partial Q}{\partial z}=\frac{\partial R}{\partial y}.$$

Эти равенства можно переписать следующим образом:

$$\frac{\partial R}{\partial y}-\frac{\partial Q}{\partial z}=0,$$

$$\frac{\partial P}{\partial z}-\frac{\partial R}{\partial x}=0,$$

$$\frac{\partial Q}{\partial x}-\frac{\partial P}{\partial y}=0.$$

Выражения, стоящие в левых частях последних трех равенств, играют важную роль в векторном анализе. Итак, в случае трех переменных мы имеем три условия полного дифференциала, в то время как для двух переменных было достаточно одного.

Если считать, что величины P, Q и R являются проекциями на оси некоторого вектора $\mathbf{E}$, то есть $\mathbf{E}=P(x, y, z)\,\mathbf{i}+Q(x, y, z)\,\mathbf{j}+R(x, y, z)\,\mathbf{k}$, где $\mathbf{i}$, $\mathbf{j}$, $\mathbf{k}$ — единичные векторы осей, то величины

$$\frac{\partial R}{\partial y}-\frac{\partial Q}{\partial z}, \qquad \frac{\partial P}{\partial z}-\frac{\partial R}{\partial x}, \qquad \frac{\partial Q}{\partial x}-\frac{\partial P}{\partial y}$$

можно рассматривать как проекции на оси некоторого другого вектора. Последний вектор называют *вихрем* или *ротором* вектора $\mathbf{E}$ и обозначают символом $\operatorname{rot}\mathbf{E}$. Понятие ротора вектора играет чрезвычайно важную роль в гидро- и электродинамике.

$$\operatorname{rot}\mathbf{E}=\left(\frac{\partial R}{\partial y}-\frac{\partial Q}{\partial z}\right)\mathbf{i}+\left(\frac{\partial P}{\partial z}-\frac{\partial R}{\partial x}\right)\mathbf{j}+\left(\frac{\partial Q}{\partial x}-\frac{\partial P}{\partial y}\right)\mathbf{k}.$$

Следовательно, условием полного дифференциала будет: $\operatorname{rot}\mathbf{E}=0$.

ПАРАМЕТРИЧЕСКОЕ ПРЕДСТАВЛЕНИЕ КРИВЫХ И ПОВЕРХНОСТЕЙ

Остановимся теперь на двух важных непрерывных функциях, геометрическое истолкование которых приводит к более общему и глубокому понятию кривой и поверхности. Областью определения первой функции *является отрезок прямой, вторая функция определена в некоторой области или интервале на плоскости.* В первом случае точки области определения (как отмечалось выше) можно рассматривать как вещественные числа, тогда мы будем иметь дело с некоторой вещественной переменной t, принимающей значения в некотором интервале: $a \leq t \leq b$. Таким образом, координаты x, y, z образов точек оказываются непрерывными функциями переменной (параметра) t:

Первая из двух упомянутых выше непрерывных функций является непрерывным образом (на плоскости или в пространстве) отрезка $[a\,;\,b]$. Множество образов точек отрезка мы называем кривой. Однако, поскольку так определенное понятие включает в себя и множества точек, которые совсем не похожи на кривую в ее наглядном понимании, то данное определение обычно сужают. Назовем поэтому простой дугой образ отрезка $[a\,;\,b]$ при таком отображении, которое разные точки области определения отображает в разные точки пространства (то есть для которого существует обратное отображение).

Заметим, что и последнее определение допускает кривые, для которых ни в одной точке не существует касательной. Геометрические фигуры, называемые кривыми из наглядных соображений, мы получим, если дополнительно потребуем, чтобы функции $x(t)$, $y(t)$ и $z(t)$ были непрерывно дифференцируемы (то есть такие дифференцируемые функции, производные которых тоже дифференцируемы). Кроме того, пусть выполняется неравенство $x'(t)^2 + y'(t)^2 + z'(t)^2 > 0$, где знак «$'$» (штрих) обозначает дифференцирование по t. Кривые, удовлетворяющие перечисленным условиям, называются *гладкими*.

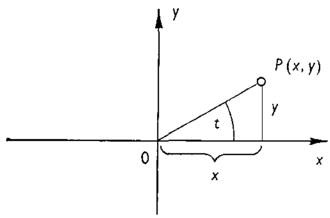
Данное определение допускает и хорошо знакомое представление кривых на плоскости в явном виде: $y = f(x)$. В самом деле, если координаты всех образов по оси z равны 0, то кривая лежит в плоскости xy. Если теперь положить $y = y(x)$, $x = t$, то мы получаем «привычную» явную кривую. Новое определение имеет еще одно преимущество: оно не делает никакого различия между координатами x и y, а рассматривает их абсолютно равноправно.

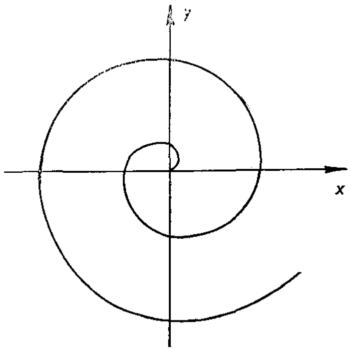
Примеры:

1. Пусть P — произвольная точка кривой. Обозначим через t угол между отрезком OP и положительной полуосью Ox. Пусть $\overline{OP} = kt(k > 0$, постоянная), то есть длина отрезка OP пропорциональна t. Тогда

$$x(t) = kt\cos t$$

$$y(t) = kt\sin t \qquad (t \leq 0)$$





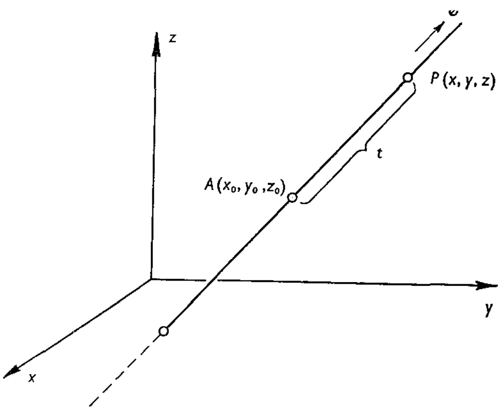
Параметр t есть угол между отрезком OP и положительной полуосью $+x$. Параметрическое представление облегчает изображение кривой

(для этой кривой обычное представление $y = f(x)$ вызвало бы явные затруднения).

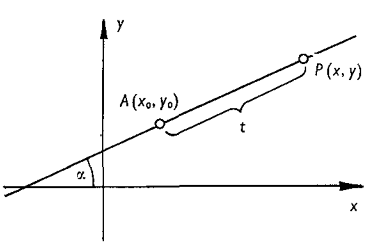
2. Рассмотрим параметрическое представление прямой в пространстве. Обозначим через t расстояние от произвольной точки $P(x; y; z)$ прямой до некоторой произвольно выбранной (но затем зафиксированной) точки $A(x_0; y_0; z_0)$ той же прямой. Установим направление на прямой, считая, что в одну сторону от точки A t принимает положительные значения, в другую — отрицательные, то есть область изменения t есть бесконечный интервал: $-\infty < t < \infty$.

Пусть **e** — единичный вектор на прямой, направленный в сторону возрастания t. Как известно, компонентами вектора **e** в прямоугольной системе координат являются направляющие косинусы прямой: $\cos\alpha$, $\cos\beta$, $\cos\gamma$, где α, β, γ — углы, образованные прямой и, соответственно, положительными полуосями x, y, z. Следующие равенства очевидны:

$$x - x_0 = t\cos\alpha$$
$$y - y_0 = t\cos\beta$$
$$z - z_0 = t\cos\gamma.$$



Параметрическое представление прямой в пространстве



Для прямой на плоскости xy $z(t) = 0$, $x(t) = x_0 + t\cos\alpha$, $y(t) = y_0 + t\sin\alpha$

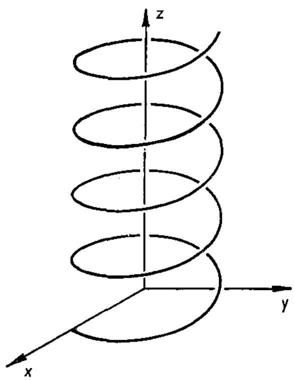


График винтовой линии

Отсюда:
$$x(t) = x_0 + t\cos\alpha$$
$$y(t) = y_0 + t\cos\beta$$
$$z(t) = z_0 + t\cos\gamma.$$

В частном случае, когда прямая лежит в плоскости xy, имеем: $z(t) \equiv 0$, $x(t) = x_0 + t\cos\alpha$, $y(t) = y_0 + t\sin\alpha$, где α — угол между прямой и положительной полуосью x.

3. В качестве третьего примера рассмотрим параметрическое представление винтовой линии:

$$x(t) = a\cos t,$$
$$y(t) = a\sin t,$$
$$z(t) = ct$$

(a и c — постоянные, $-\infty < t < \infty$).

Обратимся теперь к рассмотрению второй непрерывной функции, упомянутой в начале этой главы. Областью определения функции является некоторый интервал I на плоскости. Координаты точек на плоскости, в отличие от координат x, y, z образов точек, будем обозначать буквами u и v. Тогда упомянутый интервал можно задать неравенствами

$$a \leq u \leq b, \qquad c \leq v \leq d.$$

Следовательно, координаты x, y, z образов точек интервала есть функции двух переменных u и v:

$$x(u, v), \qquad y(u, v), \qquad z(u, v).$$

Непрерывный образ интервала I, получаемый таким образом, носит название непрерывной поверхности.

Заметим, что и здесь данное определение поверхности допускает множества точек, никак не отвечающие привычным наглядным представлениям о поверхности. По этой причине принято вводить также понятие *простого куска поверхности*, предполагая, что отображение взаимно-однозначно. Новое определение допускает и изученное ранее представление поверхности в явном виде: $z = f(x, y)$. Чтобы убедиться в этом, достаточно положить $u = x$, $v = y$ и $z(u, v) = f(x, y)$. Одно из преимуществ данного определения в том, что оно не делает различия между координатами x, y и z.

Образ F: $\{x(u, v), y(u, v), z(u, v)\}$ интервала I называется *гладким куском поверхности*, если частные производные

$$\frac{\partial x}{\partial u}, \quad \frac{\partial x}{\partial v}, \quad \frac{\partial y}{\partial u}, \quad \frac{\partial y}{\partial v}, \quad \frac{\partial z}{\partial u}, \quad \frac{\partial z}{\partial v}$$

непрерывны и выполняется неравенство

$$\begin{vmatrix} \frac{\partial y}{\partial u} & \frac{\partial z}{\partial u} \\ \frac{\partial y}{\partial v} & \frac{\partial z}{\partial v} \end{vmatrix}^2 + \begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial z}{\partial u} \\ \frac{\partial x}{\partial v} & \frac{\partial z}{\partial v} \end{vmatrix}^2 + \begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial y}{\partial u} \\ \frac{\partial x}{\partial v} & \frac{\partial y}{\partial v} \end{vmatrix}^2 > 0.$$

В каждой точке гладкого куска поверхности можно построить касательную плоскость к поверхности. Плоскость, касательная к поверхности в некоторой точке P_0 может быть определена, например, следующим образом: выберем на поверхности кроме точки P_0 две произвольные точки P_1 и P_2 так, чтобы точки P_0, P_1 и P_2 не лежали на одной прямой, и построим плоскость $P_0P_1P_2$. Если теперь устремить точки P_1 и P_2 (так, чтобы они все время оставались на поверхности) к точке P_0, то плоскость $P_0P_1P_2$ будет стремиться к некоторой предельной плоскости (не зависящей от начального выбора точек P_1 и P_2). Эта предельная плоскость и есть *касательная плоскость* к поверхности в точке P_0.

Единичный вектор, перпендикулярный к касательной плоскости в точке касания, называется *единичным вектором нормали* к поверхности и обозначается, как правило, через $\mathbf{n}$.

Из сказанного вытекает, что вектор $\mathbf{n}$ может иметь одно из двух противоположных направлений. Обычно направление вектора $\mathbf{n}$ выбирают, исходя из конкретных геометрических соображений. Поскольку направление вектора $\mathbf{n}$, очевидно, изменяется от точки к точке, то $\mathbf{n} = \mathbf{n}(x; y; z)$ или, еще проще, $\mathbf{n} = \mathbf{n}(P_0)$.

В заключение отметим, что касательная плоскость к поверхности, заданной функцией $z = f(x, y)$, в точке $(x_0; y_0; z_0 = f(x_0, y_0))$, описывается следующим уравнением:

$$\boxed{(z - z_0) = (x - x_0) \cdot f'_x(x_0, y_0) + (y - y_0) \cdot f'_y(x_0, y_0).}$$

ПРОИЗВОДНАЯ ПО НАПРАВЛЕНИЮ. ГРАДИЕНТ

Частная производная $\frac{\partial z}{\partial x}$ выражает «скорость» изменения функции $z = f(x, y)$ при постоянном y, то есть при движении в плоскости xy вдоль прямой $y = \text{const}$. Точно так же частная производная $\frac{\partial z}{\partial y}$ выражает скорость изменения функции $z = f(x, y)$ при движении вдоль прямой $x = \text{const}$. А как следует определить скорость изменения некоторой функции $z = f(x, y)$ при движении вдоль прямой, заданной уравнениями $x(t) = x_0 + t\cos\alpha$ и $y(t) = y_0 + t\sin\alpha$? Для этого необходимо рассмотреть производную сложной функции

$$z = f(x_0 + t\cos\alpha;\ y_0 + t\sin\alpha)$$

по параметру t. Эта производная называется *производной* функции $z = f(x, y)$ *по направлению* α и обозначается символом $f'_{(\alpha)}$.

$$f'_{(\alpha)} = \frac{dz}{dt} = \frac{\partial z}{\partial x}\frac{dx}{dt} + \frac{\partial z}{\partial y}\frac{dy}{dt} = \frac{\partial z}{\partial x}\cos\alpha + \frac{\partial z}{\partial y}\sin\alpha.$$

Градиент. Если принять во внимание, что $\mathbf{e} = \mathbf{i}\cos\alpha + \mathbf{j}\sin\alpha$ — единичный вектор направления α и рассмотреть вектор

$$\operatorname{grad}\mathbf{f} = \frac{\partial f}{\partial x}\mathbf{i} + \frac{\partial f}{\partial y}\mathbf{j}$$

(читается как «градиент f»), то тогда

$$f'_{(\alpha)} = \mathbf{e}_\alpha \cdot \operatorname{grad}\mathbf{f}.$$

Поскольку вектор grad **f** не зависит от направления α, а $\mathbf{e}_\alpha$ — единичный вектор, то

$$|f'_{(\alpha)}| = |\operatorname{grad}\mathbf{f}| \cdot \cos(\operatorname{grad}\mathbf{f}, \mathbf{e}_\alpha).$$

Отсюда следует, что абсолютная величина производной по направлению достигает своего максимума, когда векторы $\mathbf{e}_\alpha$ и grad **f** имеют одинаковое направление. Следовательно, наибольшее изменение функции $f(x, y)$ происходит в направлении вектора grad **f**.

Совершенно аналогично определяется производная по направлению $[\alpha;\ \beta;\ \gamma]$ функции $f(x, y, z)$ трех переменных:

$$\boxed{\begin{aligned} f'_{(\alpha)}(x, y, z) &= \frac{df(x_0 + t\cos\alpha,\ y_0 + t\cos\beta,\ z_0 + t\cos\gamma)}{dt} = \\ &= \frac{\partial f}{\partial x}\cos\alpha + \frac{\partial f}{\partial y}\cos\beta + \frac{\partial f}{\partial z}\cos\gamma. \end{aligned}}$$

Определяя, как и выше, вектор

$$\operatorname{grad}\mathbf{f} = \frac{\partial f}{\partial x}\mathbf{i} + \frac{\partial f}{\partial y}\mathbf{j} + \frac{\partial f}{\partial z}\mathbf{k}$$

(где **i**, **j**, **k** — единичные векторы координатных осей), получаем:

$$\boxed{f'_{(\alpha)}(x, y, z) = \mathbf{e}_{[\alpha;\beta;\gamma]} \cdot \operatorname{grad}\mathbf{f}(x, y, z).}$$

Здесь $\mathbf{e}_{[\alpha;\ \beta;\ \gamma]}$ — единичный вектор направления $[\alpha;\ \beta;\ \gamma]$.

КРИВОЛИНЕЙНЫЙ ИНТЕГРАЛ

Если некоторое тело под действием постоянной силы движется по прямолинейной траектории, то совершаемая при этом работа равна произведению составляющей силы по направлению траектории на длину пути, пройденного телом. Математически это можно выразить следующим образом: работа равна скалярному произведению вектора силы на вектор перемещения.

Однако это верно лишь в случае прямолинейной траектории и постоянной силы. Возникает вопрос, как вычислить совершаемую работу, если тело движется по криволинейной траектории.

Приближенное значение совершенной работы можно получить, если участок кривой от начальной точки A до конечной точки B разделить на части, и отрезки между соседними точками деления приближенно считать прямыми. Далее следует вычислить указанным выше способом величину работы на каждом отрезке разбиения, а затем сложить полученные результаты.

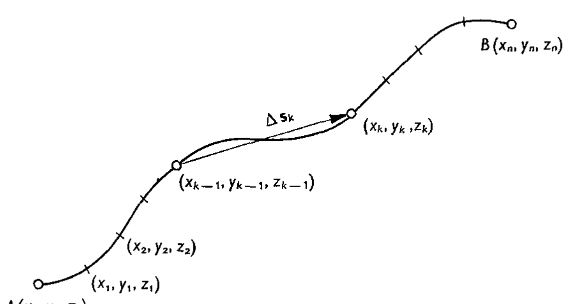
Итак, рассмотрим разбиение кривой $\overset{\frown}{AB}$ точками

$$A = (x_0;\ y_0;\ z_0), \qquad (x_1;\ y_1;\ z_1), \ldots, (x_n;\ y_n;\ z_n) = B.$$

Внутри каждой дуги разбиения (или на ее границе) выберем произвольную точку. Пусть это будут точки

$$(\xi_1;\ \eta_1,\ \zeta_1), \qquad (\xi_2;\ \eta_2;\ \zeta_2), \ldots, (\xi_n;\ \eta_n;\ \zeta_n).$$

Предположим, что вдоль k-той дуги разбиения сила сохраняет постоянное значение, равное величине силы, действующей в точке $(\xi_k;\ \eta_k;\ \zeta_k)$. Тогда приближенное значение работы, совершаемой силой $\mathbf{E}$ по передвижению тела вдоль k-той дуги разбиения, равно скалярному произведению $\mathbf{E}$ $(\xi_k;\ \eta_k;\ \zeta_k) \cdot \Delta\mathbf{s}_k$, где $\Delta\mathbf{s}_k$ обозначает вектор, соединяющий точку $(x_{k-1};\ y_{k-1};\ z_{k-1})$ с точкой $(x_k;\ y_k;\ z_k)$ (сила $\mathbf{E}$ является функцией точки).



Работа силы $\mathbf{E}$ по передвижению тела вдоль k-той дуги разбиения

Пусть

$$E_x = P(x, y, z),$$
$$E_y = Q(x, y, z),$$
$$E_z = R(x, y, z)$$

— компоненты силы $\mathbf{E}$, а Δx_k, Δy_k, Δz_k — компоненты вектора $\Delta \mathbf{s}_k$ в прямоугольной системе координат. Тогда приближенная величина работы, совершаемой на k-том участке разбиения, равна

$$\mathbf{E}(\xi_k; \eta_k; \zeta_k) \cdot \Delta \mathbf{s}_k = P(\xi_k; \eta_k; \zeta_k) \cdot \Delta x_k + Q(\xi_k; \eta_k; \zeta_k) \cdot \Delta y_k + R(\xi_k; \eta; \zeta_k;) \cdot \Delta z_k.$$

Приближенное значение полной работы по передвижению тела вдоль кривой $\overset{\frown}{AB}$ получается суммированием «элементарных работ»:

$$\sum_{k=1}^{n} \mathbf{E}(\xi_k; \eta_k; \zeta_k) \cdot \Delta \mathbf{s}_k =$$
$$= \sum_{k=1}^{n} [P(\xi_k ; \eta_k ; \zeta_k) \cdot \Delta x_k + Q(\xi_k ; \eta_k ; \zeta_k) \cdot \Delta y_k + R(\xi_k ; \eta_k ; \zeta_k) \cdot \Delta z_k].$$

Предел этой суммы (если он существует) при неограниченном измельчении разбиения кривой, то есть при неограниченном увеличении числа точек разбиения таким образом, чтобы $\max_k \Delta x_k$, $\max_k \Delta y_k$ и $\max_k \Delta z_k \to 0$, мы и примем за точную величину совершаемой работы. Этот предел обозначается как

$$\boxed{\int_{\overset{\frown}{AB}} [P(x, y, z)\, dx + Q(x, y, z)\, dy + R(x, y, z)\, dz]}$$

и называется *криволинейным интегралом*, взятым по кривой $\overset{\frown}{AB}$ от тройки функций $P(x, y, z)$, $Q(x, y, z)$, $R(x, y, z)$ (или от векторной функции $\mathbf{E}$). Величина интеграла дает значение работы, совершаемой силой $\mathbf{E}$ на пути $\overset{\frown}{AB}$. Очевидно, что в данном определении не имеет никакого значения, являются ли функции P, Q и R компонентами некоторой силы, поэтому можно сформулировать следующее общее определение:
предел, указанный выше, называется криволинейным интегралом от дифференциального выражения $P(x, y, z)\, dx + Q(x, y, z)\, dy + R(x, y, z)\, dz$, взятым по кривой $\overset{\frown}{AB}$.

Совершенно аналогично определяется криволинейный интеграл от дифференциального выражения $P(x, y)\, dx + Q(x, y)\, dy$ в случае функций двух переменных. Из самого определения ясно, что если двигаться по кривой $\overset{\frown}{AB}$ в противоположном направлении, то есть от B к A, то изменится лишь знак криволинейного интеграла.

Вычисление криволинейного интеграла. Вычисление криволинейного интеграла сводится к вычислению обыкновенного определенного интеграла. Для этого необходимо представить соответствующую кривую $\widehat{AB}$ в параметрическом виде:

$$x = x(t), \qquad y = y(t), \qquad z = z(t), \qquad a \leqq t \leqq b,$$

где $x(a)$, $y(a)$, $z(a)$ — координаты точки A, а $x(b)$, $y(b)$, $z(b)$ — координаты точки B. После этого вместо дифференциалов dx, dy, dz следует подставить, соответственно $\frac{dx}{dt}dt$, $\frac{dy}{dt}dt$, $\frac{dz}{dt}dt$, а в функциях $P(x, y, z)$, $Q(x, y, z)$ и $R(x, y, z)$ вместо переменных x, y и z следует записать их выражения: $x(t)$, $y(t)$, $z(t)$. Таким образом, мы придем к определенному интегралу

$$\int_a^b \left[P(x(t);\ y(t);\ z(t))\frac{dx}{dt} + Q(x(t);\ y(t);\ z(t))\frac{dy}{dt} + (Rx(t);\ y(t);\ z(t))\frac{dz}{dt}\right] dt,$$

который зависит только от t и значение которого совпадает со значением криволинейного интеграла. Совершенно аналогично следует поступать и в случае двух переменных.

Пример: Вычислим значение криволинейного интеграла от дифференциального выражения

$$(x + y)\,dx + xy\,dy,$$

взятого по параболе $y = x^2$ от точки $A(0;\ 0)$ до точки $B(1;\ 1)$. Задать параболу параметрически можно, например, следующим образом:

$$x = \sqrt{t}, \qquad y = t \qquad (0 \leqq t \leqq 1)$$

$$\frac{dx}{dt} = \frac{1}{2\sqrt{t}}, \qquad \frac{dy}{dt} = 1, \qquad P(x, y) = x + y, \qquad Q(x, y) = xy.$$

Следовательно,

$$\int_{\widehat{AB}} [P(x, y)\,dx + Q(x, y)\,dy] = \int_0^1 \left[(\sqrt{t} + t)\frac{1}{2\sqrt{t}} + \sqrt{t}\cdot t\cdot 1\right] dt =$$

$$= \int_0^1 \left[\frac{1}{2} + \frac{\sqrt{t}}{2} + t^{3/2}\right] dt = \left[\frac{t}{2} + \frac{1}{3}t^{3/2} + \frac{2}{3}t^{5/2}\right]_0^1 = \frac{37}{30}.$$

Условия независимости криволинейного интеграла от пути. В физике исключительную важность имеет вопрос, при каких условиях величина криволинейного интеграла не зависит от кривой (пути), соединяющей точки A и B. Можно доказать, что величина интеграла не зависит от пути из A в B (а зависит только от самих точек A и B) тогда и только тогда, когда дифференциальное выражение под интегралом представляет собой *полный дифференциал.* Действительно, если мы имеем под интегралом полный дифференциал некоторой функции $u(x, y, z)$ (представим себе опять это дифференциальное выражение как элементарную работу), то равенство

$$du = P\,dx + Q\,dy + R\,dz$$

выражает тот факт, что элементарная работа равна изменению функции $u(x, y, z)$ (так называемой *потенциальной функции*). Следовательно, полная работа на пути от A до B равна разности между значениями функции $u(x, y, z)$ в точках B и A, то есть равна $u(B) - u(A)$. Поскольку в этом случае компоненты P, Q и R вектора силы $\mathbf{E}$ равны, соответственно, $\frac{du}{dx}$, $\frac{du}{dy}$ и $\frac{du}{dz}$, то

$$\mathbf{E} = \operatorname{grad} u\,.$$

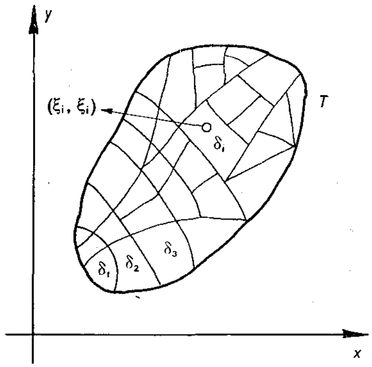
Как мы уже видели, условием существования потенциальной функции (то есть условием того, чтобы дифференциальное выражение $Pdx + Qdy + Rdz$ было полным дифференциалом некоторой функции $u(x, y, z)$) является выполнение равенства $\operatorname{rot} \mathbf{E} = 0$. Следовательно, равенство $\operatorname{rot} \mathbf{E} = 0$ является условием того, чтобы вектор $\mathbf{E}$ был градиентом некоторой функции $u(x, y, z)$: $\mathbf{E} = \operatorname{grad} u$. (В случае двух переменных условием существования потенциальной функции является выполнение только одного равенства: $\frac{\partial P}{\partial Q} - \frac{\partial y}{\partial x} = 0$).

ДВОЙНОЙ ИНТЕГРАЛ, ТРОЙНОЙ ИНТЕГРАЛ (ИНТЕГРАЛ ПО ОБЪЕМУ), ПОВЕРХНОСТНЫЙ ИНТЕГРАЛ

Двойной интеграл. Пусть $f(x, y)$ — непрерывная функция, определенная в некоторой ограниченной области T плоскости xy. Разобьем область T сетью кривых на конечное число частей, которые пронумеруем произвольным образом: $\delta_1, \delta_2, \ldots, \delta_n$. Обозначим через t_i площадь фигуры δ_i. Внутри каждой фигуры (или на ее границе) выберем произвольным образом точку $(\xi_i;\ \eta_i)$ и составим произведения $f(\xi_i;\ \eta_i)\cdot t_i$, которые затем просуммируем:

$$\sum_{i=1}^{n} f(\xi_i;\ \eta_i)\cdot t_i\,.$$

Если над каждой из фигур δ_i построить цилиндр с высотой $f(\xi_i;\ \eta_i)$, то объем такого цилиндра будет равен $f(\xi_i;\ \eta_i)\cdot t_i$, при этом записанная выше сумма даст нам общий объем всех цилиндров. Нетрудно видеть, что эта сумма есть не что иное, как приближенное значение объема части пространства, заключенной между поверхностью $z = f(x, y)$ и областью T (вспомним геометрический смысл интегральных сумм для определенного интеграла функции одной переменной). Будем теперь неограниченно измельчать разбиение, то есть неограниченно увеличивать число фигур δ_i таким образом, чтобы наибольший из диаметров фигур δ_i стремился к 0 (диаметром множества H называется $\sup d(P_1, P_2)$, где $P_1, P_2 \in H$). Можно доказать, что при этом указанные выше суммы, независимо от способа разбиения области T, сходятся к некоторому пределу, который называется *двойным* интегралом функции $f(x, y)$ в области T и обозначается символом



Определение приближенного значения объема части пространства, заключенной между поверхностью $z = f(x, y)$ и областью T

$$\boxed{\iint\limits_{(T)} f(x, y)\, dx\, dy.}$$

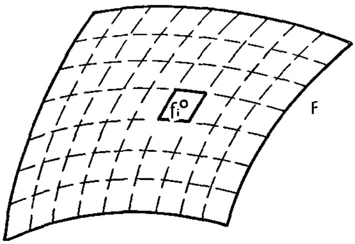
Очевидно, что (в случае $f(x, y) \geq 0$) значение этого интеграла равно объему части пространства, заключенной между поверхностью $z = f(x, y)$ и областью T.

Тройной интеграл. Совершенно аналогично определяется и так называемый *тройной интеграл* (или *интеграл по объему*) от вещественной непрерывной функции $u(x, y, z)$ трех переменных в области T пространства. Аналогично и его обозначение:

$$\boxed{\iiint\limits_{(T)} f(x, y, z)\, dx\, dy\, dz.}$$

Заметим, что тройной интеграл не имеет такого наглядного геометрического смысла, как определенный интеграл функции одной переменной или двойной интеграл.

Мы не станем здесь останавливаться на вычислении двойного и тройного интегралов, отметим лишь, что в случае сравнительно простых областей T вычисление двойного интеграла сводится к вычислению двух, а тройного — к вычислению трех обыкновенных определенных интегралов.



Разбиение поверхности F сетью кривых на конечное число кусков

Поверхностный интеграл. В теоретической физике важную роль играет и так называемый *поверхностный интеграл.*

Пусть $\Phi(x, y, z)$ — вещественная непрерывная функция, определенная в некоторой области T пространства, и пусть в той же области T параметрические уравнения

$$x = x(u, v),$$
$$y = y(u, v),$$
$$z = z(u, v)$$

задают некоторый гладкий кусок поверхности F (здесь u, v — независимые переменные (параметры), принимающие значения в некотором интервале $a \leq u \leq b$, $c \leq v \leq d$ на плоскости).

Разобьем поверхность F сетью кривых на конечное число кусков $f_1, f_2, \ldots, f_n$. Их площади обозначим, соответственно, через $t_1, t_2, \ldots, t_n$. (Если F — простой кусок поверхности, то, поскольку он является образом интервала $a \leq u \leq b$, $c \leq v \leq d$ на плоскости, указанное разбиение F порождает и разбиение интервала на некоторые подобласти $\delta_1, \delta_2, \ldots, \delta_n$).

Выберем на каждом из кусков разбиения f_i произвольным образом точку $(\xi_i; \eta_i; \zeta_i)$ и рассмотрим произведения

$$\Phi(\xi_i; \eta_i; \zeta_i) \cdot t_i,$$

а также их сумму:

$$\sum_{i=1}^{n} \Phi(\xi_i; \eta_i; \zeta_i) \cdot t_i.$$

Если теперь неограниченно измельчать разбиение поверхности F, то есть неограниченно увеличивать число кусков f_i таким образом, чтобы диаметры всех кусков стремились к 0, то можно доказать, что при этом указанные выше суммы, независимо от способа разбиения поверхности F, стремятся к некоторому пределу. Этот предел обозначается символом

$$\boxed{\int\limits_{(F)} \Phi(x, y, z)\, df}$$

и называется *поверхностным интегралом* от функции $\Phi(x, y, z)$ по поверхности F.

Заметим лишь, что вычисление этого интеграла обычно сводится к вычислению двойного интеграла в интервале $a \leq u \leq b$, $c \leq v \leq d$ на плоскости. Точнее, имеет место следующее соотношение:

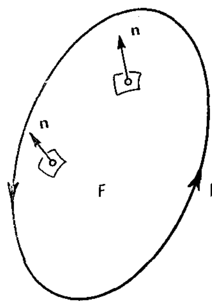
$$\int\limits_{(F)} \Phi(x, y, z) df = \iint\limits_{(I)} \Phi\big(x(u, v), y(u, v), z(u, v)\big) \sqrt{EG - F^2}\, du\, dv.$$

Здесь (I) обозначает интервал $a \leqq u \leqq b$, $c \leqq v \leqq d$ на плоскости, а E, F и G (присутствующую здесь величину F не следует путать с поверхностью F) означают следующее:

$$\boxed{\begin{aligned} E &= \left(\frac{\partial x}{\partial u}\right)^2 + \left(\frac{\partial y}{\partial u}\right)^2 + \left(\frac{\partial z}{\partial u}\right)^2 \\ F &= \frac{\partial x}{\partial u}\frac{\partial x}{\partial v} + \frac{\partial y}{\partial u}\frac{\partial y}{\partial v} + \frac{\partial z}{\partial u}\frac{\partial z}{\partial v} \\ G &= \left(\frac{\partial x}{\partial v}\right)^2 + \left(\frac{\partial y}{\partial v}\right)^2 + \left(\frac{\partial z}{\partial v}\right)^2 \end{aligned}}$$

(E, F и G — так называемые *гауссовы коэффициенты* поверхности. См. гл. «*Дифференциальная геометрия*», стр. 357).

Теорема Стокса. Теорема Стокса устанавливает важную для теоретической физики связь между поверхностными и криволинейными интегралами.



Из каждой точки поверхности F восстановим единичный вектор нормали

Пусть в пространстве задана векторная функция, то есть каждой точке пространства поставлен в соответствие некоторый вектор $\mathbf{E}$. С такими векторными функциями можно встретиться, например, при рассмотрении силовых полей, когда в каждой точке пространства на помещенную туда массу действует некоторая сила, зависящая, вообще говоря, от координат точки: $\mathbf{E} = \mathbf{E}(x, y, z)$. Компоненты вектора $\mathbf{E}$ по осям x, y и z обозначим, соответственно, через $\mathbf{P}(x, y, z)$, $\mathbf{Q}(x, y, z)$ и $\mathbf{R}(x, y, z)$. Пусть в этом векторном пространстве (то есть в той части пространства, где определен вектор $\mathbf{E}$) задана простая замкнутая кривая l. Установим направление на кривой l и рассмотрим некоторый гладкий кусок поверхности F, для которого кривая l является границей. Из каждой точки поверхности F восстановим единичный вектор нормали таким образом, чтобы круговое направление на кривой l и направление вектора нормали соответствовали бы направлению вращения и направлению продвижения при завинчивании винта (с правой езьбой).

Рассмотрим теперь поверхностный интеграл

$$\int\limits_{(F)} \operatorname{rot}\mathbf{E} \cdot \mathbf{n}\, df,$$

где под интегралом записано скалярное произведение векторов $\operatorname{rot}\mathbf{E}$ и $\mathbf{n}$. Рассмотрим также криволинейный интеграл по кривой l (в заданном направлении):

$$\int_l \mathbf{P}(x, y, z)dx + \mathbf{Q}(x, y, z)dy + \mathbf{R}(x, y, z)dz] .$$

Теорема Стокса утверждает, что эти интегралы равны:

$$\int_{(F)} \operatorname{rot} \mathbf{E} \cdot \mathbf{n}\, df = \int_l [\mathbf{P}(x, y, z)dx + \mathbf{Q}(x, y, z)dy + \mathbf{R}(x, y, z)dz] .$$

Интегральная теорема Гаусса—Остроградского. Эту теорему также часто используют физики-теоретики.

Дивергенцией вектора $\mathbf{E}(x, y, z)$ (обозначается $\operatorname{div} \mathbf{E}$) называется скалярная величина

$$\boxed{\operatorname{div} \mathbf{E} = \frac{\partial P}{\partial x} + \frac{\partial Q}{\partial y} + \frac{\partial R}{\partial z} ,}$$

где, как обычно,

$$P(x, y, z) ,$$

$$Q(x, y, z) ,$$

$$R(x, y, z)$$

обозначают компоненты вектора $\mathbf{E}$ по осям x, y, z. Рассмотрим теперь простую гладкую замкнутую поверхность F в пространстве и обозначим ограниченную ею область пространства через V. Пусть $\mathbf{n}$ — единичный вектор нормали к поверхности F, направленный во внешнюю сторону. Рассмотрим поверхностный интеграл

$$\int_{(F)} \mathbf{E} \cdot \mathbf{n}\, df ,$$

где $\mathbf{E} \cdot \mathbf{n}$ обозначает скалярное произведение векторов. Рассмотрим также тройной интеграл

$$\iiint_{(V)} \operatorname{div} \mathbf{E}\, dx\, dy\, dz .$$

Теорема Гаусса—Остроградского утверждает, что эти интегралы равны:

$$\boxed{\int_{(F)} \mathbf{E} \cdot \mathbf{n}\, df = \iiint_{(V)} \operatorname{div} \mathbf{E}\, dx\, dy\, dz.}$$

НЕЯВНЫЕ ФУНКЦИИ

Уже в элементарной аналитической геометрии мы встречаемся с кривыми, которые заданы не в явном виде $y = f(x)$ или $x = \varphi(y)$, а представлены уравнением более общего вида: $F(x, y) = 0$. Таковы, например, уравнения окружности: $x^2+y^2-r^2=0$, или эллипса: $\frac{x^2}{a^2}+\frac{y^2}{b^2}-1=0$. Чтобы выразить y как функцию x (или x как функцию y), такое уравнение надо решить. Говорят, что получаемая таким образом функция $y=f(x)$ (или $x=\varphi(y)$) задана уравнением $F(x,y)=0$ в *неявном виде* и только после решения уравнения может быть представлена в *явном виде*: $y=f(x)$ (или $x=\varphi(x)$).

В приведенных выше примерах, а также во многих случаях, встречающихся на практике, явное решение может быть легко выражено через известные элементарные функции. Нередко, однако, решение удается выразить лишь с помощью бесконечных рядов, в таких случаях искомая функция может быть найдена приближенно, с любой степенью точности.

Исследование уравнения $F(x,y)=0$. Однако значительно проще во многих отношениях не решать уравнение $F(x,y)=0$, а провести исследование рассматриваемой функции непосредственно на уравнении $F(x,y)=0$.

Заметим сразу, что было бы ошибкой полагать, будто всякой функции $F(x,y)$ уравнение $F(x,y)=0$ действительно ставит в соответствие некоторую функцию $y=f(x)$ (или $x=\varphi(y)$). Это видно на примере уравнения $x^2+y^2+1=0$ или уравнения $x^2+y^2=0$, которому удовлетворяет лишь одна пара значений $x=0$ и $y=0$.

По этой причине необходимо прежде всего выяснить, когда уравнение $F(x,y)=0$ действительно задает в неявном виде некоторую функцию $y=f(x)$, а затем исследовать свойства этой функции.

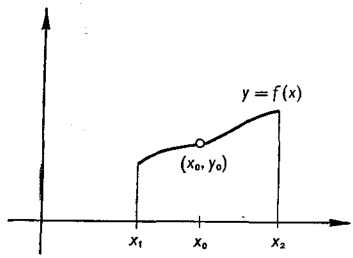
Ситуация станет понятней, если функцию $z=F(x,y)$ изобразить в виде поверхности в пространстве. Пересечем эту поверхность плоскостью xy и исследуем вопрос, существует ли кривая, являющаяся линией их пересечения и представимая в виде $y=f(x)$ (или $x=\varphi(y)$). Из геометрических соображений сразу ясно, когда такая кривая может существовать, а когда нет. Пусть $F(x_0,y_0)=0$, то есть $(x_0;\ y_0)$ — общая точка поверхности с плоскостью xy. Если плоскость xy касается поверхности в точке $(x_0;\ y_0)$, и в окрестности точки $(x_0;\ y_0)$ не имеет с поверхностью других общих точек, то кривой пересечения не существует. Если же плоскость xy не является касательной к поверхности, то, очевидно, существует кривая их пересечения. Поскольку в точке $(x_0;y_0)$ $z_0=F(x_0,y_0)=0$, а уравнение касательной плоскости к поверхности в этой точке имеет вид

$$z=(x-x_0)F'_x(x_0,y_0)+(y-y_0)F'_y(x_0,y_0),$$

то требование, чтобы плоскость xy не была касательной к поверхности, означает, что $F'_x(x_0,y_0)$ и $F'_y(x_0,y_0)$ не должны одновременно обращаться в нуль. Если предположить, что частные производные непрерывны, то по крайней мере одна из них отлична от нуля в некоторой окрестности точки $(x_0;\ y_0)$. Поскольку переменные x и y рассматриваются абсолютно равноправно, то без ограничения общности можно предположить, что $F'_y\neq 0$.

Сформулируем теперь общую теорему существования неявных функций, дающую одновременно и правило их дифференцирования.

Пусть частные производные F'_x и F'_y функции $F(x,y)$ непрерывны, и пусть в некоторой точке $(x_0;\ y_0)F(x_0,y_0)=0$, и $F'_y(x_0,y_0)\neq 0$. Тогда



Уравнение $F(x, y)=0$ задает некоторую однозначную непрерывную функцию $y=f(x)$

существует окрестность $x_1 \leq x \leq x_2$ точки x_0 (на оси x), в которой уравнение $F(x, y) = 0$ задает некоторую однозначную непрерывную функцию $y = f(x)$, причем $y_0 = f(x_0)$. Во всех точках x интервала (x_1, x_2) выполняется равенство $F(x, f(x)) = 0$. Функция $y = f(x)$ дифференцируема в интервале (x_1, x_2) и ее производная

$$y' = f'(x) = -\frac{F'_x(x, f(x))}{F'_y(x, f(x))} .$$

В большинстве случаев это правило позволяет дифференцировать неявную функцию значительно проще, чем если выразить ее в явном виде. Например,

$$F(x, y) = x^2 + y^2 - 1 = 0 ,$$

$$y' = -\frac{F'_x}{F'_x} = -\frac{2x}{2y} = -\frac{x}{y} .$$

Неявные функции нескольких переменных. Изложенное выше можно распространить на функции нескольких переменных.

Пусть $u = F(x, y, z)$ — функция трех переменных. В предположениях, аналогичных сделанным выше, и при условии $F'_z \neq 0$ уравнение $F(x, y, z) = 0$ задает некоторую функцию

$$z = f(x, y) ,$$

дифференцируемую по x и по y. Частные производные $\frac{\partial z}{\partial x} = z'_x$ и $\frac{\partial z}{\partial y} = z'_y$ могут быть получены из следующих равенств:

$$F'_x + F'_z \frac{\partial z}{\partial x} = 0 \quad (y \text{ — фиксированная постоянная}),$$

$$F'_y + F'_x \frac{\partial z}{\partial x} = 0 \quad (x \text{ — фиксированная постоянная}).$$

Отсюда:

$$\boxed{\frac{\partial z}{\partial x} = -\frac{F'_x}{F'_z} , \quad \frac{\partial z}{\partial y} = -\frac{F'_y}{F'_z} .}$$

ФУНКЦИИ n ПЕРЕМЕННЫХ

В заключение остановимся на функциях с числом переменных, большим трех. В этом случае геометрическая иллюстрация независимых переменных возможна лишь в абстрактном пространстве, размерность которого больше трех. И хотя такие пространства являются лишь математическими построениями, ввести их имеет смысл, так как они во многом аналогичны трехмерному пространству, что послужило источником немалого числа важных открытий.

Назовем n-мерным евклидовым пространством упорядоченное множество всевозможных строк из n вещественных чисел вида $(x_1, x_2, \ldots, x_n)$. Причем каждую такую строку будем называть точкой пространства и под расстоянием между двумя точками $X(x_1, x_2, \ldots, x_n)$ и $Y(y_1, y_2, \ldots, y_n)$ будем понимать число

$$d(X, Y) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \ldots + (x_n - y_n)^2}.$$

Очевидно, что такое расстояние удовлетворяет хорошо известным аксиомам:

1. $d(X; Y) = 0$ тогда и только тогда, когда $X = Y$;
2. $d(X; Y) = d(Y; X)$;
3. если $Z(z_1, z_2, \ldots, z_n)$ — некоторая третья точка пространства, то $d(X; Y) + d(Z; Z) \geqq d(X; Z)$ (неравенство треугольника).

Понятия сходимости последовательности точек, предельной точки множества, окрестности точки, понятия открытого множества, области и другие также можно распространить на n-мерные пространства. Рассматривая отображения таких пространств друг в друга, нетрудно получить обобщение понятия функции. Функцию n переменных (скалярную) мы получим, если рассмотрим отображение некоторого подмножества H n-мерного пространства на вещественную числовую ось.

ДИФФЕРЕНЦИАЛЬНЫЕ УРАВНЕНИЯ

Уравнение, в котором неизвестной величиной является функция одной или нескольких переменных, и в котором, кроме самой функции, присутствуют и ее производные, мы называем дифференциальным уравнением. Если неизвестная функция зависит от нескольких переменных, то мы говорим об уравнении с частными производными, если же это функция одной переменной, то имеем обыкновенное дифференциальное уравнение. Согласно тому, какой порядок имеет самая высокая производная, говорят об уравнениях 1-го, 2-го, ..., n-го порядка, то есть порядок дифференциального уравнения равен порядку высшей производной, присутствующей в уравнении.

Примеры: (y — неизвестная функция, x — независимая переменная).

1. $$y' - y \sin x = \sqrt{y} + 1$$

— обыкновенное дифференциальное уравнение первого порядка.

2. $$y'' \cos x - \frac{y'}{y} \ln x = y^2$$

— обыкновенное дифференциальное уравнение второго порядка.

3. $$\frac{\partial^2 u}{\partial x^2} = \frac{\partial^2 u}{\partial x \partial y} + u^2 \frac{\partial u}{\partial x} = \left(\frac{\partial u}{\partial y}\right)^2$$

— уравнение с частными производными второго порядка (неизвестная функция $u(x, y)$ — функция переменных x и y).

Решить дифференциальное уравнение или, другими словами, проинтегрировать его означает найти все те функции, которые удовлетворяют данному уравнению.

Мы изложим здесь простейшие сведения об обыкновенных дифференциальных уравнениях. Рассмотрим сначала несколько примеров из физики.

1. На тело, движущееся по прямой, в направлении движения действует некоторая постоянная сила. Как зависит путь, пройденный телом, от времени?

Обозначим длину пути через s, а время — через t. Постоянная сила вызывает постоянное ускорение, которое обозначим через g. Поскольку ускорение есть вторая производная пути по времени, то получаем следующее дифференциальное уравнение:

$$\frac{d^2 s}{dt^2} = g = \text{const}.$$

Это обыкновенное дифференциальное уравнение второго порядка. Поскольку

$$\frac{d^2 s}{dt^2} = \frac{d}{dt}\left(\frac{ds}{dt}\right) = g,$$

то

$$\frac{ds}{dt} = gt + c_1,$$

где c_1 — произвольная постоянная. Следовательно, скорость тела $\frac{ds}{dt}$ есть линейная функция времени. Если в начальный момент времени ($t = 0$) скорость тела равнялась 0, то

$$\frac{ds}{dt} - c_1 = 0,$$

в противном случае c_1 — величина начальной скорости $\left(\text{в момент времени } t = 0 \ \frac{ds}{dt} = c_1\right)$. После вторичного интегрирования получаем:

$$s = s(t) = \frac{gt^2}{2} + c_1 t + c_2,$$

где c_2 — произвольная постоянная. Физический смысл постоянной c_2 — путь, пройденный телом до момента времени $t = 0$. Если пройденный путь отмерять от того места, где тело находилось в момент $t = 0$, то $c_2 = 0$. Если и начальная скорость (c_1) равна 0, то уравнение движения принимает вид $s = \frac{gt^2}{2}$ (уравнение свободного падения). Важно отметить, что в общем случае уравнение движения содержит две независимые друг от друга произвольные постоянные: c_1 и c_2. Это характерно для дифференциальных уравнений второго порядка.

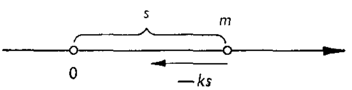
2. На точку с массой m, движущуюся вдоль прямой, действует сила, направленная в сторону фиксированной точки 0 на прямой и по величине пропорциональная расстоянию массы от точки 0. Имеем:

$$m\frac{d^2s}{dt^2} = -ks,$$

т. е.

$$\frac{d^2s}{dt^2} + \frac{k}{m}\cdot s = 0.$$

Можно доказать, что любое решение этого дифференциального уравнения второго порядка представляется в виде $s = s(t) = \varrho \sin\left(\sqrt{\frac{k}{m}}t + \varphi\right)$, где ϱ и φ — произвольные постоянные, причем всегда можно считать, что $\varrho \geq 0$. Это известное уравнение гармонического колебания. Таким образом точка с массой m совершает вокруг точки 0 гармоническое колебание с частотой



Точка с массой m совершает гармоническое колебание вокруг точки 0

$$n = \frac{1}{2\pi}\sqrt{\frac{k}{m}}.$$

3. Если в предыдущем примере на точку с массой m кроме рассмотренной «упругой» силы действует еще сила сопротивления среды, по величине пропорциональная скорости $\frac{ds}{dt}$ точки (и направленная, естественно, в противоположную сторону), то уравнение движения принимает следующий

вид:

$$m\frac{d^2s}{dt^2} = -l\frac{ds}{dt} - ks$$

или

$$\frac{d^2s}{dt^2} + \frac{l}{m}\frac{ds}{dt} + \frac{k}{m}s = 0,$$

где l — постоянная сопротивления среды. Это уравнение может иметь три различные решения:

I. $\frac{4k}{m} - \left(\frac{l}{m}\right)^2 > 0$ (сила сопротивления среды здесь меньше упругой силы). Здесь «сила сопротивления среды» и «упругая сила» есть не локальные значения $l\frac{ds}{dt}$ и ks и не величины $\frac{4k}{m}$ и $\left(\frac{l}{m}\right)^2$, а некоторые качественные характеристики, интерпретирующие это дифференциальное уравнение (прим. ред.).

II. $\frac{4k}{m} - \left(\frac{l}{m}\right)^2 < 0$ (сила сопротивления среды превышает упругую силу)

III. $\frac{4k}{m} - \left(\frac{l}{m}\right)^2 = 0$ (сила сопротивления среды превышает упругую силу).

В I случае решение имеет вид:

$$s(t) = \varrho e^{-\frac{l}{2m}\cdot t} \cdot \sin\left(\sqrt{\frac{k}{m} - \left(\frac{l}{2m}\right)^2} \cdot t + \varphi\right),$$

где ϱ и φ — произвольные числа (затухающее колебание). Во II случае

$$s(t) = e^{-\frac{l}{2m}\cdot t}, \left(c_1 e^{+\sqrt{\left(\frac{l}{2m}\right)^2 - \frac{k}{m}}\cdot t} + c_2 e^{-\sqrt{\left(\frac{l}{2m}\right)^2 - \frac{k}{m}}\cdot t}\right).$$

В случае III

$$s(t) = e^{-\frac{l}{2m}\cdot t}(c_1 t + c_2),$$

где c_1 и c_2 — произвольные постоянные. В случаях II и III из-за большого сопротивления среды колебания не происходит.

ОБЫКНОВЕННОЕ ДИФФЕРЕНЦИАЛЬНОЕ УРАВНЕНИЕ ПЕРВОГО ПОРЯДКА

Общий вид обыкновенного дифференциального уравнения первого порядка:

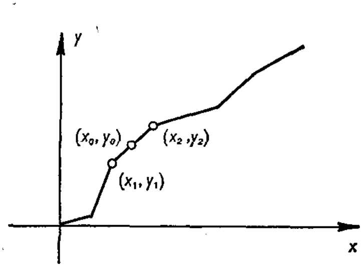
$$F(y', y, x) = 0 .$$

Если это уравнение можно разрешить относительно производной y', то есть привести его к виду

$$y' = f(x, y) ,$$

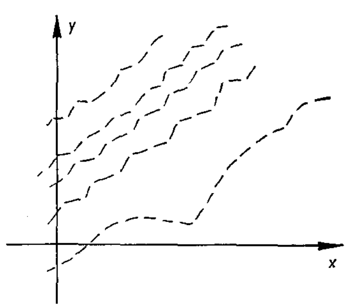
то получаем дифференциальное уравнение, заданное в *явной форме*, в противном случае имеем уравнение, заданное в *неявной форме*. Мы рассмотрим здесь уравнения, заданные в явной форме.

Поле направлений. Уравнению $y' = f(x, y)$ на координатной плоскости xy можно придать наглядный смысл с помощью так называемого поля направлений. Для этого в каждой точке $(x;\ y)$ плоскости установим направление с угловым коэффициентом $f(x, y)$. Для искомой функции $y = y(x)$ характерно то, что угловой коэффициент касательной к кривой $y = y(x)$ в любой ее точке $(x;\ y(x))$ равен $f(x, y)$. Таким образом, кривая решения точно «согласуется» с полем направлений: направление кривой в любой ее точке совпадает с направлением поля в той же точке. Зная это, можно с любой степенью точности находить решения $y = y(x)$ графически.



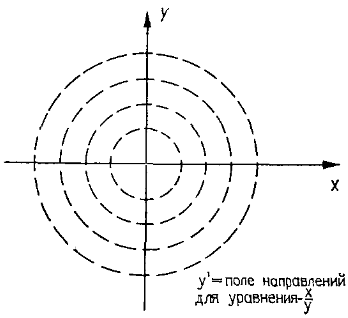
Приближенное изображение решения $y = y(x)$

Выберем на плоскости xy произвольную точку $(x_0;\ y_0)$, в которой функция $f(x, y)$ определена, и проведем через эту точку небольшой отрезок прямой с угловым коэффициентом $f(x_0, y_0)$. Из концов $(x_1;\ y_1)$ и $(x_2;\ y_2)$ полученного отрезка проведем небольшие отрезки прямых с угловыми коэффициентами, соответственно, $f(x_1, y_1)$ и $f(x_2, y_2)$. Продолжая этот процесс дальше в обе стороны, мы получим некоторую ломаную линию, приближенно изображающую решение $y = y(x)$. Чем короче отрезки ломаной, тем точнее приближение.



Маленькие отрезки сливаются в сплошные кривые

Если в качестве начальной точки вместо $(x_0;\ y_0)$ взять некоторую другую точку, то мы получим некоторую другую ломаную линию, являющуюся приближением некоторого другого решения уравнения $y' = f(x, y)$. Отсюда видно, что существует бесконечно много решений уравнения (так как в качестве начальной точки можно выбрать любую точку из области определения функ-



Интегральные кривые представляют собой концентрические окружности с центром в начале координат

ции $f(x, y)$. Более общую картину мы получим, если изобразим поле направлений, соответствующее уравнению $y' = f(x, y)$. Для этого из возможно большего числа точек $(x; y)$ (в принципе это следовало бы сделать для каждой точки области определения функции) проведем короткие отрезки прямых с угловыми коэффициентами $f(x, y)$. Если выполнить это для достаточно большого количества точек, то построенные отрезки сольются в сплошные кривые, изображающие решения дифференциального уравнения и называемые *интегральными кривыми*. В качестве примера нарисуем поле направлений для дифференциального уравнения $y' = -\frac{x}{y}$. На рисунке хорошо видно, что интегральные кривые для этого уравнения представляют собой концентрические окружности с центром в начале координат. Нетрудно заметить также, что через каждую точку плоскости, за исключением начала координат, проходит ровно одна интегральная кривая.

РЕШЕНИЕ УРАВНЕНИЯ $y' = f(x, y)$ МЕТОДОМ ПОСЛЕДОВАТЕЛЬНЫХ ПРИБЛИЖЕНИЙ

После всего сказанного о поле направлений можно подумать, что через каждую точку области определения функции $f(x, y)$ (за исключением некоторых точек, называемых *особыми)* проходит одна и только одна интегральная кривая. Чаще всего (и в подавляющем большинстве случаев, представляющих интерес для практики) так оно и есть. Сейчас мы познакомимся с процессом, позволяющим с любой степенью точности находить решение уравнения $y' = f(x, y)$, проходящее через данную точку $(x_0; y_0)$.

Метод последовательных приближений. Назовем 0-вым приближенным решением функцию $y_0(x) = y_0 = \text{const}$. 1-ое приближенное решение определим следующим образом:

$$y_1(x) = y_0 + \int_{x_0}^{x} f(x, y_0)dx .$$

В предположении, что $y_n(x)$ уже определено, положим

$$y_{n+1}(x) = y_0 + \int_{x_0}^{x} f(x, y_n(x)dx .$$

Таким образом, каждому натуральному числу n мы поставили в соответствие некоторую функцию $y_n(x)$. Можно доказать, что при довольно общих предположениях относительно функции $f(x, y)$ последовательность функций $y_0(x), y_1(x), \ldots, y_n(x), \ldots$ равномерно сходится к некоторой функции $y(x)$, которая удовлетворяет уравнению

$$y(x) = y_0 + \int_{x_0}^{x} f(x, y(x))dx,$$

то есть

$$y'(x) = f(x, y(x)) .$$

Причем, если $(f'(x, y)| \leq M$, то при $n > M \mid x - x_0 \mid$ $|y(x) - y_{n+1}(x) \mid \leq$ $\leq \frac{(x - x_0)^{n+1} M^{n+1}}{n!}$

Следовательно, функция $y(x)$ оказывается решением дифференциального уравнения $y' = f(x, y)$ и удовлетворяет начальному условию $y(x_0) = y_0$, то есть проходит через точку $(x_0 ; y_0)$. Итак, метод последовательных приближений при довольно общих предположениях доказывает существование решения дифференциального уравнения $y' = f(x, y)$ (решения, удовлетворяющего начальному условию $y(x_0 = y_0)$. Такую теорему называют *теоремой существования*. Кроме того, метод последовательных приближений дает простой и практичный способ нахождения приближенных решений дифференциального уравнения с любой степенью точности.

Проиллюстрируем использование этого метода на примере уравнения $y' = y$. Начальным условием пусть будет $y_0(0) = 1$, то есть $x_0 = 0, y_0 = 1$. В данном случае $f(x, y) = y$. В качестве последовательных приближений получаем следующие функции:

$$y_0(x) = 1,$$

$$y_1(x) = 1 + \int_0^x 1 \cdot dx = 1 + x,$$

$$y_2(x) = 1 + \int_0^x (1 + x)dx = 1 + x + \frac{x^2}{2},$$

$$y_3(x) = 1 + \int_0^x \left((1+x) + \frac{x^2}{2}\right) dx = 1+x+ \frac{x^2}{2} + \frac{x^3}{2 \cdot 3},$$

$$\vdots$$

$$y(x) = 1 + \int_0^x \left(1+x+ \frac{x^2}{2} + \frac{x^3}{2 \cdot 3} + \ldots + \frac{x^{n+1}}{(n-1)!}\right) dx =$$

$$= 1+x+\frac{x^2}{2}+\frac{x^3}{3!}+\ldots+\frac{x^n}{n!}.$$

$$\vdots$$

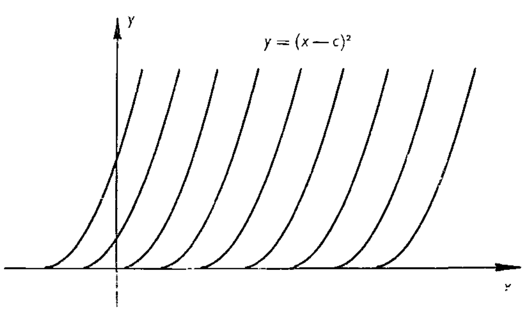
Нетрудно видеть, что $y_n(x) \to \sum_{n=0}^{\infty}\frac{x^n}{n!} = e^x$.

Нам остается лишь точнее сформулировать, что понимается под «довольно общими предположениями» относительно функции $f(x, y)$. К ним относится, например, предположение об ограниченности функции $f(x, y)$ и ее частной производной $f'_y(x, y)$ в некоторой области на плоскости, то есть условия

$$|f(x, y)| \leqq M_1; \qquad |f'_y(x, y)| \leqq M_2.$$

Если функция $f(x, y)$ не удовлетворяет этим условиям, то может случиться, что, например, через одну точку плоскости проходит несколько интегральных кривых. Примером может служить дифференциальное уравнение $y = 2\sqrt{y}$, где рассматривается положительное значение корня. В этом случае функция $f(x, y) = +2\sqrt{y}$ определена для всех положительных y. Во всякой ограниченной полосе на плоскости имеем: $|f(x, y)| \leqq M_1$. Однако, если приближаться к оси x, то $f'_y(x, y) = \frac{1}{\sqrt{y}}$ неограниченно возрастает. Нетрудно убедиться в том, что интегральными кривыми в этом случае являются правые ветви парабол $y = (x - c)^2$, где c — произвольная положительная постоянная (левые ветви парабол являются интегральными кривыми для дифференциального уравнения $y = -2\sqrt{y}$). Однако и сама ось x, поскольку функция $y \equiv 0$ удовлетворяет уравнению, тоже является интегральной кривой, которая «огибает» параболы. Функцию $y \equiv 0$ мы называем особым решением дифференциального уравнения.

Введем еще несколько терминов, употребляемых в связи с решениями дифференциальных уравнений.



Интегральными кривыми являются правые ветви парабол $y = (x - c)^2$, а также и сама ось x, представляющая собой особое решение

Общим решением дифференциального уравнения называется всякая функция $y(x, c)$ (представимая, может быть, лишь в неявном виде $\varphi(x, y, c) = 0$), содержащая произвольную постоянную c и удовлетворяющая дифференциальному уравнению.

Если в общее решение подставить какое-либо конкретное значение c, то мы получим некоторое частное решение дифференциального уравнения. Таким образом общее решение можно рассматривать как совокупность всех частных решений.

Семейство интегральных кривых, соответствующих частным решениям, дает общее решение уравнения. Огибающая этого семейства кривых представляет собой *особое решение*, которое не входит в общее решение дифференциального уравнения.

РЕШЕНИЕ ОБЫКНОВЕННОГО ДИФФЕРЕНЦИАЛЬНОГО УРАВНЕНИЯ ПЕРВОГО ПОРЯДКА С ПОМОЩЬЮ РАЗДЕЛЕНИЯ ПЕРЕМЕННЫХ

Если правая часть уравнения $y' = f(x, y)$ имеет вид $f(x, y) = g(x)\,h(y)$, то есть представляет собой произведение двух функций, одна из которых зависит только от x, а другая — только от y, то тогда

$$\frac{1}{h(y)} \cdot \frac{dy}{dx} = g(x), \quad \int \frac{1}{h(y)} \frac{dy}{dx}\,dx = \int g(x)dx\,.$$

Производя замену переменных под интегралом в левой части последнего равенства, получаем:

$$\int \frac{1}{h(y)}\,dy, \quad \text{то есть}$$

$$\int \frac{dy}{h(y)} = \int g(x)dx\,.$$

Решение исходного уравнения свелось к вычислению двух неопределенных интегралов. После интегрирования мы получим обыкновенное (не дифференциальное) уравнение относительно y.

Итак, в случае уравнений вида $y' = g(x)\,h(y)$ формально следует поступать следующим образом: умножив обе части на dx, перенести его в правую часть уравнения, а $h(y)$ перенести в знаменатель левой части, после чего проинтегрировать левую часть уравнения по y, а правую — по x.

Примеры: $y' = -\dfrac{x}{y}$. Имеем:

$$y\,dy = -x\,dx, \quad \int y\,dy = -\int x\,dx,$$

$$\frac{y^2}{2} = -\frac{x^2}{2} + C,$$

$$x^2 + y^2 = 2C\,.$$

Мы получили уравнение окружности, что совпадает с результатом графического решения, приведенного на стр. 506. Рассмотрим другой пример:

$y' = -ky$ (k — постоянная).

$$\frac{dy}{y} = -k\,dx, \quad \int \frac{dy}{y} = -kx + C,$$

$$\ln y = -kx + C,\ y = e^c \cdot e^{-kx},\ \text{если } x = 0,\ y_0 = e^c,\ \text{то есть } y = y_0 e^{-kx}.$$

В уравнении $y' = x + y$ переменные не разделяются. Однако, если произвести замену $u = x + y$, $u' = 1 + y'$, то мы получим уравнение $u' = u + 1$ с разделяющимися переменными.

ОБЫКНОВЕННОЕ ДИФФЕРЕНЦИАЛЬНОЕ УРАВНЕНИЕ ВТОРОГО ПОРЯДКА

Остановимся кратко на дифференциальном уравнении второго порядка, имеющем вид $y'' = f(x, y, y')$. Здесь также может быть использован метод последовательных приближений. С этой целью введем переменную $z = y'$, тогда исходное уравнение сведется к системе двух дифференциальных уравнений первого порядка:

$$y' = z$$
$$z' = f(x, y, z).$$

В этом случае также можно задать начальные условия: $y(x_0) = y_0$, $z(x_0) = z_0$. Это равносильно заданию двух начальных условий для исходного уравнения: $y(x_0) = y_0$, и $y'(x_0) = y_0' = z_0$.

Пусть теперь

$$\begin{aligned}
&y_0(x) = y_0 = \text{const} && z_0(x) = z_0 = \text{const}\\
&y_1(x) = y_0 + \int_{x_0}^{x} z_0(x)dx && z_1(x) = z_0 + \int_{x_0}^{x} f(x, y_0, z_0)dx\\
&y_2(x) = y_0 + \int_{x_0}^{x} z_1(x)dx && z_2(x) = z_0 + \int_{x_0}^{x} f\big(x, y_1(x), z_1(x)\big)\,dx\\
&y_3(x) = y_0 + \int_{x_0}^{x} z_2(x)\,dx && z_3(x) = z_0 + \int_{x_0}^{x} f(x, y_2(x), z_2(x))dx\\
&\vdots\\
&y_n(x) = y_0 + \int_{x_0}^{x} z_{n-1}(x)dx && z_n(x) = \int_{x_0}^{x} f\big(x, y_{n-1}(x), z_{n-1}(x)\big)dx\\
&\vdots
\end{aligned}$$

В этом случае также можно доказать, что при довольно общих предположениях относительно функции $f(x, y, y')$ . . . последовательности функций

$y_0(x), y_1(x), \ldots, y_n(x), \ldots$ и $z_0(x), z_1(x), \ldots, z_n(x), \ldots$ равномерно сходятся, соответственно, к некоторым функциям $y(x)$ и $z(x) = y'(x)$, которые, очевидно, удовлетворяют начальным условиям $y(x_0) = y_0$ и $z(x_0) = y'(x_0) = z_0 = y_0'$, кроме того, $y'(x) = z(x)$, $z'(x) = y''(x) = f(x, y, z) = f(x, y, y')$.

Как и в случае уравнения первого порядка, применение метода последовательных приближений с одной стороны доказывает существование решения *дифференциального уравнения* $y'' = f(x, y, y')$, *удовлетворяющего* начальным условиям $y(x_0) = y_0$ и $y'(x_0) = y_0'$, с другой стороны, дает возможность найти это решение с любой степенью точности. Кроме того, выясняется, что при довольно общих условиях, накладываемых на функцию $f(x, y, y')$ существует только одно решение уравнения, удовлетворяющее заданным начальным условиям (это называют единственностью решения дифференциального уравнения).

Общим решением дифференциального уравнения второго порядка является всякая функция $y(x, c_1, c_2)$ (представимая, быть может, и в неявном виде: $\varphi(x, y, c_1, c_2) = 0$), которая удовлетворяет уравнению и содержит два произвольных постоянных параметра, принимающих независимо друг от друга вещественные числовые значения. При подстановке в общее решение конкретных числовых значений постоянных c_1 и c_2 получается частное решение уравнения. Задавая два начальных условия, мы тем самым определяем значения постоянных c_1 и c_2, получая при этом частное решение уравнения, удовлетворяющее данным начальным условиям. Уравнение второго порядка также может иметь особые решения, которые не входят в общее решение.

В заключение отметим, что, аналогично случаю уравнения второго порядка, метод последовательных приближений можно использовать и в случае уравнения n-го порядка, заданного в явной форме:

$$y^{(n)} = f(x, y', y'' \ldots, y^{(n-1)}) .$$

В этом случае заменой переменных

$$y' = z_1,\ y'' = z_2 = z_1',\ \ldots,\ y^{(n-1)} = z_{n-1} = z_{n-2}',$$
$$y^{(n)} = z_n = f(x, y, z_1, \ldots, z_{n-1})$$

исходное уравнение приводится к системе n дифференциальных уравнений первого порядка.

РЕШЕНИЕ ДИФФЕРЕНЦИАЛЬНЫХ УРАВНЕНИЙ С ПОМОЩЬЮ СТЕПЕННЫХ РЯДОВ

Встречаются случаи, когда метод последовательных приближений оказывается громоздким и неудобным. В таких случаях довольно часто может быть использован метод, заключающийся в следующем: предположим, что искомое решение в некотором интервале представимо в виде суммы сходящегося степенного ряда, то есть

$$y(x) = a_0 + a_1x + a_2x^2 + \ldots + a_n x^n + \ldots,$$

где коэффициенты ряда пока неизвестны. Подставим эту функцию (дифференцируя ряд почленно) в дифференциальное уравнение. Полученное уравнение приведем к виду, где в правой части стоит 0, а в левой — ряд, расположенный по возрастающим степеням x. Тогда все коэффициенты ряда в левой части равны 0. Из получаемых таким образом равенств можно по порядку вычислять коэффициенты исходного ряда.

Проиллюстрируем использование этого метода на примере уравнения $y'' = -y$.

Итак, пусть

$$y = a_0 + a_1x + a_2x^2 + \ldots + a_n x^n + \ldots$$

Тогда

$$y' = a_1 + 2a_2x + 3a_3x^2 + \ldots + na_n x^{n-1} + \ldots$$

$$y'' = 2a_2 + 2 \cdot 3a_3x + 3 \cdot 4a_4x^2 + \ldots + n(n-1)a_nx^{n-2} + \ldots$$

Подставляя полученное выражение для y'' в уравнение $y'' + y = 0$, и располагая левую часть по возрастающим степеням x, получим:

$$(a_0 + 2a_2) + (a_1 + 2 \cdot 3 \cdot a_3)x + (a_2 + 3 \cdot 4 \cdot a_4)x^2 + \ldots + \\ + (a_n + (n+1)(n+2)a_{n+2})x^n + \ldots = 0\,.$$

Отсюда:

$$\begin{aligned} &a_0 + 2a_2 = 0 \\ &a_1 + 2 \cdot 3 \cdot a_3 = 0 \\ &a_2 + 3 \cdot 4 \cdot a_4 = 0 \\ &\ldots\ldots\ldots\ldots\ldots\ldots \\ &\ldots\ldots\ldots\ldots\ldots\ldots \\ &a_n + (n+1)\,(n+2)a_{n+2} = 0 \\ &\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ &\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \end{aligned}$$

Таким образом, коэффициенты исходного ряда еще не определены, так как мы знаем лишь, что

$$a_2 = -\frac{a_0}{2}$$

$$a_3 = -\frac{a_1}{2 \cdot 3}$$

$$a_4 = -\frac{a_2}{3 \cdot 4} = \frac{a_0}{4!}$$

$$a_5 = -\frac{a_3}{4 \cdot 5} = \frac{a_1}{5!}$$

. . . .
. . . .

$$a_{n+2} = -\frac{a_n}{(n+1)(n+2)}$$

. . . .
. . . .

Нетрудно заметить, что каждый коэффициент выражается через другой коэффициент, индекс которого на два меньше. Следовательно, если, например, задать значение a_0, то тем самым будут определены все коэффициенты с четными индексами. Точно так же задание значения a_1 определяет все коэффициенты с нечетными индексами. Отсюда следует, что для полного определения степенного ряда необходимо и достаточно задать значения коэффициентов a_0 и a_1. Но это в точности соответствует тому, что для определения частного решения дифференциального уравнения второго порядка следует задать два начальных условия:

$$a_0 = y(0), \qquad a_1 = y'(0) \,.$$

Пусть, например, $a_0 = 1$, $a_1 = 0$, тогда все коэффициенты с нечетными индексами равны 0, а для коэффициентов с четными индексами имеем:

$$a_1 = 1, \; a_2 = -\frac{1}{2}, \; a_4 = \frac{1}{4!}, \; a_6 = -\frac{1}{6!}, a_8 = \frac{1}{8!}, \ldots$$

Следовательно,

$$y(x) = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} + \frac{x^6}{6!} + \ldots = \cos x \,.$$

В случае $a_0 = 0$, $a_1 = 1$ аналогичным образом получим разложение в степенной ряд для $\sin x$.

ЛИНЕЙНЫЕ ДИФФЕРЕНЦИАЛЬНЫЕ УРАВНЕНИЯ

Уравнения, левая часть которых представляет собой линейное выражение относительно функции y и ее производных, называются линейными дифференциальными уравнениями и имеют в случае порядка n вид:

$$y^{(n)} + a_{n-1}(x)y^{(n-1)} + a_{n-2}(x)y^{(n-2)} + \ldots + a_1(x)y' + a_0(x)y = b(x) \,,$$

где коэффициенты $a_k(x)$ $(k = 0, 1, \ldots, n-1)$ и $b(x)$ — известные функции. Если $b(x) \equiv 0$, то получаем однородное линейное уравнение, в противном случае — неоднородное.

Решение однородного уравнения. В случае уравнения первого порядка, то есть уравнения вида

$$y' + p(x)y = q(x),$$

решение находится интегрированием. Однородное уравнение $y' + p(x)y = 0$ — с разделяющимися переменными:

$$\frac{dy}{y} = -p(x)\,dx, \int \frac{dy}{y} = \ln n = -\int p(x)\,dx + c$$

$$y(x) = e^c \cdot e^{-\int p(x)\,dx} = c \cdot e^{-\int p(x)\,dx} \quad (c = e^c).$$

Решение неоднородного уравнения будем искать в виде

$$y = C(x) \cdot e^{-\int p(x)\,dx},$$

то есть подставив в общее решение однородного уравнения вместо посто. янной C неизвестную функцию. Имеем:

$$y' = C'(x)e^{-\int p(x)\,dx} - C(x)\,p(x)\,e^{-\int p(x)\,dx}.$$

После подстановки в неоднородное уравнение получаем:

$$e^{-\int p(x)\,dx}\{C'(x) - C(x)\,p(x) + p(x)\,C(x)\} = q(x)$$

и, следовательно,

$$C'(x) = q(x)\,e^{\int p(x)\,dx}$$

Отсюда:

$$C(x) = \int q(x)\,e^{\int p(x)\,dx}\,dx + K,$$

где K — постоянная, появляющаяся при последнем интегрировании. Следовательно, общее решение неоднородного уравнения

$$y(x) = e^{-\int p(x)\,dx}\left\{\int q(x)e^{\int p(x)\,dx}\,dx + K\right\}.$$

Общий вид линейного дифференциального уравнения второго порядка:

$$\boxed{y'' + a_1(x)y' + a_0(x)y = b(x).}$$

Это уравнение играет важную роль в физике. Частный случай этого уравнения, когда $a_1(x)$ и $a_2(x)$ — постоянные, а $b(x) \equiv 0$ мы уже рассматривали в начале главы как уравнение затухающих колебаний. Для однородного уравнения, то есть когда $b(x) \equiv 0$, имеет место следующая теорема: если известны два нетривиальные, независимые друг от друга решения $y_1(x)$ и $y_2(x)$ однородного уравнения, то общее решение дается формулой $C_1y_1(x) + C_2y_2(x)$, где C_1 и C_2 — произвольные постоянные (тривиальным решением называется функция $y(x) \equiv 0$; функции $y_1(x)$ и $y_2(x)$ называются независимыми, если их отношение не равно постоянной). Тот факт, что вместе с функциями $y_1(x)$ и $y_2(x)$ и любая функция вида $C_1y_1(x) + C_2y_2(x)$ является решением однородного уравнения, очевидным образом вытекает из правил дифференцирования. Интересно здесь то, что, кроме функций вида $C_1y_1(x) + C_2y_2(x)$, других решений нет.

Если известно некоторое нетривиальное решение y_1 однородного уравнения, то простым интегрированием можно найти другое решение y_2, независимое от первого. В самом деле, будем искать второе решение в виде $y_2 = y_1 \int z\,dx$, где $z = z(x)$ — неизвестная функция. Подставляя это y_2 в уравнение, получаем:

$$\underbrace{\{y_1'' + a_1(x)y_1' + a_2(x)y_1\}}_{=\,0,\ \text{так как}\ y_1\ -\ \text{решение.}} \int z\,dx + z'y_1 + z(2y_1' + a_1(x)y_1) = 0$$

Отсюда, после деления на y_1, получаем:

$$z' + z\left(2\frac{y_1'}{y_1} + a_1(x)\right) = 0\cdot$$

Таким образом мы получили однородное уравнение первого порядка относительно z, его решение:

$$z = C \cdot e^{-2\int \frac{y_1'}{y_1}dx} \cdot e^{-\int a_1(x)\,dx} = C\frac{e^{-\int a_1(x)\,dx}}{y_1^2}\,.$$

Здесь можно положить $C = 1$, так как для y_2 достаточно найти одну функцию $z(x)$. Окончательно получаем:

$$y_2(x) = y_1(x)\int \frac{e^{-\int a_1(x)\,dx}}{y_1^2}\,dx.$$

Отыскание одного нетривиального решения y_1 нередко оказывается сложной задачей и подчас удается лишь с помощью степенного ряда.

Что касается неоднородного уравнения второго порядка, то, зная одно его частное решение $Y(x)$, общее решение можно получить в виде

$$Y(x) + C_1y_1(x) + C_2y_2(x)\,,$$

где $C_1y_1(x) + C_2y_2(x)$ — общее решение соответствующего однородного уравнения. Действительно, если $y(x)$ — произвольное решение неоднородного уравнения, то

$$y'' + a_1(x)y' + a_0(x)y = b(x)$$
$$Y'' + a_1(x)Y' + a_0(x)Y = b(x)\,.$$

Вычитая второе уравнение из первого, получаем:

$$(y - Y)'' + a_1(x)(y - Y)' + a_0(x)(y - Y) = 0,$$

то есть функция $y - Y$ удовлетворяет однородному уравнению, а следовательно, представима в виде $C_1y_1 + C_2y_2$.

ТЕОРИЯ ФУНКЦИЙ КОМПЛЕКСНОГО ПЕРЕМЕННОГО

МНОЖЕСТВА ТОЧЕК КОМПЛЕКСНОЙ ПЛОСКОСТИ

Аналогично теории точечных множеств на прямой, плоскости или в пространстве, теория точечных множеств комплексной плоскости может основываться на понятии расстояния. В дальнейшем мы будем отождествлять понятия «точка комплексной плоскости» и «комплексное число». Основные понятия и теоремы для комплексной плоскости формулируются точно так же, как и для вещественной прямой, поэтому мы лишь кратко перечислим их.

Под рассточнием между двумя комплексными числами понимается вещественное число

$$d(z_1, z_2) = |z_1 - z_2|\,.$$

Очевидно, что выполняются аксиомы расстояния:

1. $d(z_1, z_2) = 0$ тогда и только тогда, когда $z_1 = z_2$
2. $d(z_1, z_2) = d(z_2, z_1)$
3. $d(z_1, z_3) \leq d(z_1, z_2) + d(z_2, z_3)$ — неравенство треугольника.

ε-окрестностью комплексного числа z_0 называется множество всех комплексных чисел, которые удалены от z_0 меньше, чем на ε, то есть для которых выполняется неравенство

$$|z - z_0| < \varepsilon\,.$$

z_0 называется предельной точкой множества H, если в любой окрестности z_0 содержится бесконечно много точек из H.

Множество H называется ограниченным, если множество, состоящее из абсолютных величин его элементов, ограничено. Это означает, что все элементы ограниченного множества содержатся в круге достаточно большого радиуса с центром в нуле.

На комплексной плоскости справедлива теорема Больцано—Вейерштрасса.

Множество H называется открытым, если оно содержит всякую свою точку вместе с некоторой ее окрестностью, и замкнутым, если оно содержит все свои предельные точки.

В теории функций комплексного переменного очень важную роль играют области, то есть открытые связные множества.

Открытое множество называется связным, если любые две его точки могут быть соединены ломаной, целиком принадлежащей множеству.

Замкнутой областью мы называем множество, которое получается добавлением к области всех ее предельных точек.

Понятия и теоремы, связанные с последовательностями, также могут быть без труда перенесены на случай комплексных чисел. В связи с этим следует заметить, что

$$\lim_{n\to\infty} z_n = z_0$$

равносильно тому, что

$$\boxed{\lim_{n\to\infty} \operatorname{Re} z_n = \operatorname{Re} z_0, \qquad \lim_{n\to\infty} \operatorname{Im} z = \operatorname{Im} z_0.}$$

Аналогично положение и в случае бесконечных рядов. Необходимо, однако, отметить, что в случае комплексных степенных рядов роль интервала сходимости выполняет круг сходимости. Точнее, областью сходимости степенного ряда

$$a_0 + a_1(z - z_0) + \ldots + a_n(z - z_0)^n + \ldots = \sum_{n=0}^{\infty} a_n(z - z_0)^n$$

является круг с радиусом

$$\boxed{r = \frac{1}{\overline{\lim\limits_{n\to\infty}} \sqrt[n]{|a_n|}}.}$$

Отсюда следует, например, что геометрический ряд $1 + z + z^2 + z^3 + \ldots$ внутри единичного круга сходится, и его сумма равна $\dfrac{1}{1-z}$.

В любом круге, центр которого совпадает с центром круга сходимости, а радиус меньше радиуса сходимости, степенной ряд сходится абсолютно и равномерно, кроме того, его можно почленно дифференцировать.

ФУНКЦИИ КОМПЛЕКСНОГО ПЕРЕМЕННОГО

Пусть

$$w = f(z)$$

— комплексная функция. Это означает, что областью ее определения является некоторое множество H комплексной плоскости, а область значений лежит в некоторой другой комплексной плоскости. Если $z = x + iv$, $w = u + iv$, то функция может быть записана в виде:

$$w = f(x + iy) = u(x;\ y) + iv(x;\ y)\,.$$

Функция $f(z)$ называется непрерывной в точке z_0 множества H, если для любого положительного числа ε найдется такая окрестность точки z_0, что для всех точек множества H, попадающих в эту окрестность, выполняется неравенство

$$|f(z) - f(z_0)| < \varepsilon\,.$$

Данное определение эквивалентно следующему: функция $f(z)$ называется непрерывной в точке z_0 множества H, если для любой последовательности z_n точек множества H из $z_n \to z_0$ следует $f(z_n) \to f(z_0)$. Очевидно, что непрерывность функции $f(z)$ в точке $z_0 = x_0 + iy_0$ равносильна непрерывности функций $u(x, y)$ и $v(x, y)$ (как функций вещественных переменных) в точке $(x_0;\ y_0)$.

Функция $f(z)$ называется дифференцируемой в точке z_0, если существует

$$\lim_{z \to z_0} \frac{f(z) - f(z_0)}{z - z_0} = f'(z_0)\,.$$

Это определение эквивалентно тому, что для всякой последовательности $z_n \to z_0$ выражение

$$\frac{f(z_n) - f(z_0)}{z_n - z_0}$$

стремится к тому же пределу $f'(z_0)$. Очевидно, что если $f(z)$ дифференцируема в некоторой точке, то в этой точке функция непрерывна, в то же время из непрерывности функции не следует ее дифференцируемость. Функция, дифференцируемая во всех точках некоторой области, называется регулярной (или аналитической) в этой области. Правила дифференцирования здесь аналогичны правилам дифференцирования вещественных функций.

Примеры:

1. $f(z) = z^n\,,\qquad f'(z) = nz^{n-1}$

2. $f(z) = \dfrac{1}{z^2 + 1}\,,\qquad f'(z) = -\dfrac{2z}{(z^2 + 1)^2}\,.$

Рассмотрим примеры непрерывных, но не дифференцируемых функций.

1. $f(z) = \operatorname{Re} z$, $\quad \dfrac{\operatorname{Re}(z+h) - \operatorname{Re} z}{h} = \begin{cases} 1, & \text{если } h \text{ вещественное,} \\ 0, & \text{если } h \text{ мнимое,} \end{cases}$

2. $f(z) = \bar{z}$ (число, сопряженное с z).

В этом случае $\dfrac{\overline{z+h} - \bar{z}}{h} = \begin{cases} 1, & \text{если } h \text{ вещественное,} \\ -1, & \text{если } h \text{ мнимое.} \end{cases}$

ДИФФЕРЕНЦИАЛЬНЫЕ УРАВНЕНИЯ КОШИ—РИМАНА (ДАЛАМБЕРА—ЭЙЛЕРА), УРАВНЕНИЕ ЛАПЛАСА

Пусть $f(z)$ — дифференцируемая функция. Посмотрим, какими свойствами обладают в этом случае функции $u(x, y)$ и $v(x, y)$ (вещественная и мнимая части функции $f(z)$). Если

$$\frac{f(z+h) - f(z)}{h} \to f'(z),$$

когда h произвольным образом стремится к 0, то рассмотрим сначала случай вещественного, а затем — чисто мнимого h. В первом случае

$$\frac{f(x+h) - f(z)}{h} = \frac{u(x+h; y) - u(x; y)}{h} + i\,\frac{v(x+h; y) - v(x; y)}{h}.$$

Устремив h к 0, получаем:

$$f'(z) = \frac{\partial u}{\partial x} + i\,\frac{\partial v}{\partial x}.$$

Аналогичным образом во втором случае:

$$f'(z) = -i\,\frac{\partial u}{\partial y} + \frac{\partial v}{\partial y}.$$

Приравнивая правые части обоих равенств, получим:

$$\frac{\partial u}{\partial x} = \frac{\partial v}{\partial y}; \qquad \frac{\partial u}{\partial y} = -\frac{\partial v}{\partial x},$$

Это и есть так называемые уравнения Коши—Римана или Даламбера—Эйлера.

Итак, из дифференцируемости функции $f(z)$ следует, что ее вещественная и мнимая части имеют частные производные первого порядка, удовлетворяющие уравнениям Коши—Римана.

Обратное утверждение, вообще говоря, неверно: из справедливости уравнений Коши—Римана не следует дифференцируемость функции $f(z)$.

Однако, если предположить, что во всех точках некоторой области существуют и непрерывны частные производные первого порядка от функций u и v, то выполнение условий Коши—Римана оказывается не только необходимым, но и достаточным условием регулярности (дифференцируемости в области) функции $f(z)$.

В дальнейшем мы увидим, что дифференцируемость функции комплексного переменного оказывается значительно более сильным условием, чем дифференцируемость в вещественном случае. В самом деле, можно доказать, что если функция $f(z)$ регулярна в некоторой области (то есть дифференцируема в каждой точке области), то в этой области она бесконечно дифференцируема. Отсюда следует, что для функций $u(x, y)$ и $v(x, y)$ существуют непрерывные частные производные любого порядка. Дифференцируя уравнения Коши—Римана, получаем:

$$\frac{\partial^2 u}{\partial x^2} = \frac{\partial^2 v}{\partial x\,\partial y} = \frac{\partial^2 v}{\partial x\,\partial y} = -\frac{\partial^2 u}{\partial y^2}; \qquad \frac{\partial^2 v}{\partial x^2} = -\frac{\partial^2 v}{\partial y^2}.$$

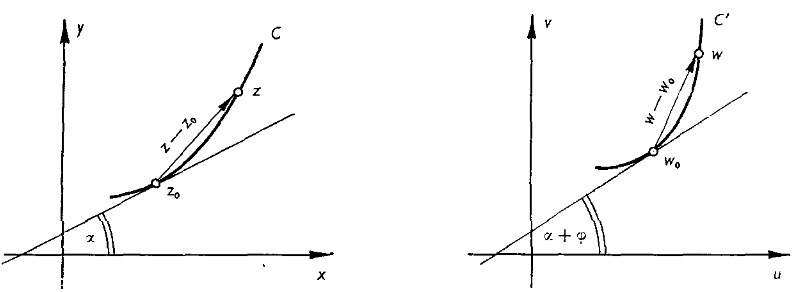
Следовательно, вещественная и мнимая части функции, регулярной в некоторой области, удовлетворяют уравнениям

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = 0 \quad \text{и} \quad \frac{\partial^2 v}{\partial x^2} + \frac{\partial^2 v}{\partial y^2} = 0.$$

Уравнения такого типа называются *уравнениями Лапласа*. Функции, удовлетворяющие уравнению Лапласа, называются гармоническими функциями. Таковы, например, функции $x^2 + y^2$; $e^x \cos y$; $\log(x^2 + y^2)$.

КОНФОРМНЫЕ ОТОБРАЖЕНИЯ (ОТОБРАЖЕНИЯ, ЗАДАВАЕМЫЕ РЕГУЛЯРНЫМИ ФУНКЦИЯМИ)

Пусть $w = f(z)$ — функция, регулярная в некоторой области T плоскости z. Рассмотрим отображение области T комплексной плоскости z на комплексную плоскость w, задаваемое этой функцией. Действительно,



$f'(z_0) \neq 0$ Точка $w = f(z)$ описывает некоторую кривую функция $f(z)$ каждой точке z области T ставит в соответствие некоторое комплексное число w, которое можно изобразить точкой на некоторой другой комплексной плоскости w. Таким образом, функция $w = f(z)$ отображает точку z в точку w. Поскольку функция отображает каждую точку z области T, то при этом вся область T как множество точек плоскости z отображается в некоторое множество T' точек плоскости w. Множество T' называется образом области T при отображении f. Выберем такую точку z_0 из T, в которой $f'(z_0) \neq 0$. Устремим теперь переменную точку z области T к точке z_0 по некоторой кривой C, касательная к которой в точке z_0 образует с осью x угол α. При этом переменная точка $w = f(z)$ опишет в плоскости w некоторую кривую C', проходящую через точку $w_0 = f(z_0)$. Поскольку

$$\arg(z - z_0) \to \alpha$$

и

$$\arg(w - w_0) - \arg(z - z_0) = \arg \frac{w - w_0}{z - z_0} \to \arg f'(z_0) = \varphi,$$

то

$$\arg(w - w_0) \to \alpha + \varphi,$$

т. е. для кривой C' определена касательная в точке w_0, образующая с вещественной осью плоскости w угол $\alpha + \varphi$. Следовательно, при отображении кривая C «повернется» на угол φ. Поскольку этот угол зависит только от точки z_0, то то же самое верно для любой кривой, проходящей через точку z_0. Отсюда следует, что при отображении $w = f(z)$ сохраняется угол между произвольными кривыми, выходящими из точки z_0, то есть образы кривых в плоскости w пересекаются под такими же углами. Такое отображение, сохраняющее углы, называют *конформным* (оно сохраняет также и ориентацию углов). Другое важное свойство конформных отображений состоит в следующем:

$$\left| \frac{w - w_0}{z - z_0} \right| \to |f'(z_0)| = k > 0,$$

то есть если z находится в достаточной близости от z_0, то отношение $|w - w_0| : |z - z_0|$ приближенно равно k. Другими словами, в достаточно малой окрестности точки z_0 отображение представляет собой растяжение с коэффициентом k.

Условие $f'(z_0) \neq 0$ существенное, в чем можно убедиться на примере отображения, задаваемого функцией $w = z^2$. В точке $z = 0$ это отображение удваивает углы.

ПРИМЕР КОНФОРМНОГО ОТОБРАЖЕНИЯ

Чтобы получить геометрическую иллюстрацию конформного отображения, задаваемого некоторой регулярной функцией $w = f(z)$, лучше всего выбрать на плоскости z подходящее семейство кривых и рассмотреть их образы в плоскости w. Если кривые семейства пересекаются, то на плоскости w они перейдут в кривые, пересекающиеся под таким же углом, и тем самым дадут наглядную картину отображения. В качестве примера рассмотрим отображение, задаваемое функцией $w = e^z$. В связи с этим мы упомянем несколько понятий, представляющих и самостоятельный интерес. Определим сначала функцию e^z комплексного переменного. Для вещественных x известно разложение в степенной ряд

$$e^x = 1 + \frac{x}{1!} + \frac{x^2}{2!} + \ldots + \frac{x^n}{n!} + \ldots$$

Обобщая это разложение на случай произвольного комплексного z, получаем:

$$e^z = 1 + \frac{z}{1!} + \frac{z^2}{2!} + \ldots + \frac{z^n}{n!} + \ldots$$

Этот степенной ряд сходится на всей комплексной плоскости. Здесь остаются в силе правила возведения в степень, имеющие место для абсолютно сходящихся рядов, таким образом, для любых комплексных чисел z_1 и z_2

$$e^{z_1+z_2} = e^{z_1} \cdot e^{z_2}.$$

В случае $z = ix$, гда x — вещественное число, отсюда можно вывести одно интересное соотношение. Имеем:

$$e^{ix} = 1 + ix + \frac{i^2x^2}{2!} + \frac{i^3x^3}{3!} + \frac{i^4x^4}{4!} + \frac{i^5x^5}{5!} + \frac{i^6x^6}{6!} + \ldots$$

Поскольку

$$\begin{array}{llll} i^2 = -1, & i^4 = 1, & i^6 = -1, & i^8 = 1, \ldots \\ i^3 = -i, & i^5 = i, & i^7 = -i, & i^9 = i, \ldots \end{array}$$

то получаем, что

$$e^{ix} = \left(1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \frac{x^6}{6!} + \ldots\right) + i\left(x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \ldots\right)$$

Первый ряд в скобках — ряд Тейлора для $\cos x$, второй — для $\sin x$. В результате получаем так называемую формулу Эйлера:

$$\boxed{e^{ix} = \cos x + i \sin x.}$$

Если $x = 2\pi n$, то $e^{i\,2\pi n} = \cos 2\pi n + i \cdot \sin 2\pi n = 1$, где n — произвольное целое число. Из формулы Эйлера следует, что выражение $r(\cos\varphi + i \sin\varphi)$, представляющее собой тригонометрическую форму комплексного числа, может быть записано в виде $z = re^{i\varphi}$, который называется экспоненциальной формой комплексного числа.

ЛОГАРИФМ КОМПЛЕКСНОГО ЧИСЛА

Под логарифмом (натуральным) комплексного числа z понимается множество всех тех комплексных чисел w, для которых

$$z = e^w.$$

Чтобы для данного z найти значения $w = \ln z$, запишем z в экспоненциальной форме $z = re^{i\varphi}$, а искомое w — в виде $w = u + iv$. Имеем:

$$e^{u+iv} = re^{i\varphi},$$

то есть

$$e^u \cdot e^{iv} = re^{i\varphi}.$$

Отсюда: $e^u = r$, $v = \varphi + 2\pi n$, где n — произвольное целое число, так как $e^{i2\pi n} = 1$. Таким образом приходим к следующему результату:

$$\boxed{\ln z = \ln|z| + i\varphi + i \cdot 2\pi n,}$$

где $\ln|z| = \ln r = u$ — единственный вещественный логарифм положительного числа $r = |z|$. Следовательно, вещественная часть логарифма комплексного числа определена однозначно, в то время как мнимая часть принимает бесконечное число значений, которые отличаются друг от друга на число, кратное $2\pi i$.

Пример: $\ln(-1) = \ln|-1| + i\pi + 2\pi in$,

Поскольку

$$-1 = 1 \cdot e^{i\pi}$$

то

$$\ln(-1) = i\pi + i \cdot 2\pi n.$$

ФУНКЦИЯ e^z

Рассмотрим теперь отображение, задаваемое функцией

$$w = e^z.$$

Запишем независимую переменную z в виде $z = x + iy$, а w — в виде $w = u + iv$. Имеем:

$$u + iv = e^{x+iy} = e^x(\cos y + i \sin y).$$

Вещественные и мнимые части, соответственно, равны, поэтому

$$u = e^x \cos y, \qquad v = e^x \sin y.$$

Отсюда получаем:

$$u^2 + v^2 = e^{2x},$$

$$\frac{v}{u} = \operatorname{tg} y.$$

Рассмотрим теперь на плоскости xy (где x — вещественная, а y — мнимая ось) прямые $x = \text{const}$ и $y = \text{const}$ и найдем их образы на плоскости uv. Прямые $x = \text{const}$ перпендикулярны оси x, прямые $y = \text{const}$ — оси y. Два эти семейства прямых пересекаются под прямым углом, следовательно, их образы при конформном отображении также будут перпендикулярны друг к другу. Если $x = \text{const}$, то

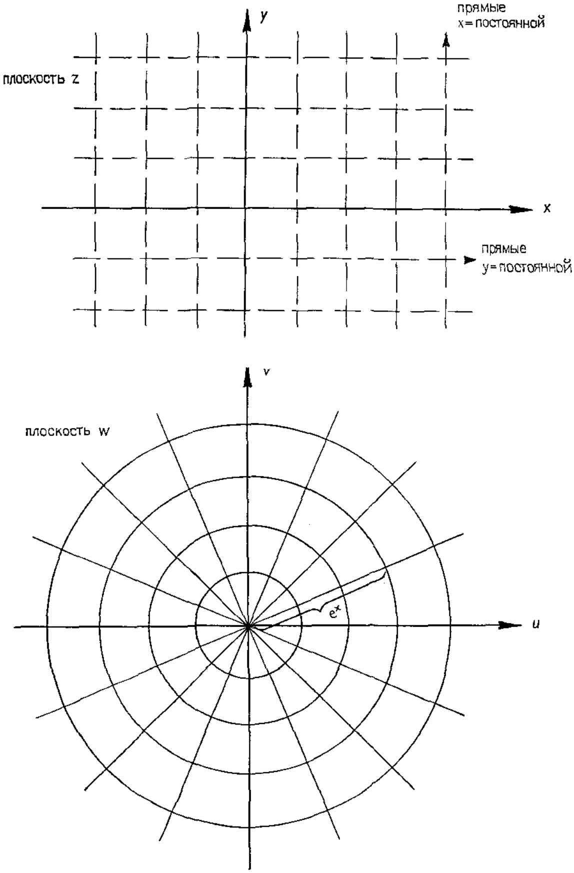
$$u^2 + v^2 = e^{2x} = \text{const},$$

то есть на плоскости uv образами прямых $x = \text{const}$ являются окружности радиуса e^x с центром в начале координат. Если $y = \text{const}$, то

$$v = (\operatorname{tg} y)u.$$

На плоскости uv это уравнение прямой, проходящей через начало координат с угловым коэффициентом $\operatorname{tg} y$. Подведем итог: образами прямых, перпендикулярных оси x, являются окружности с центром в начале координат, а образами прямых, перпендикулярных оси y, — прямые, проходящие через начало координат.

На рисунке приведена геометрическая иллюстрация рассмотренного отображения.



Образами прямых, перпендикулярных оси y, являются прямые, проходящие через начало координат, а образами прямых, параллельных оси x, — концентрические окружности с центром в начале координат

КРИВОЛИНЕЙНЫЙ ИНТЕГРАЛ ОТ ФУНКЦИИ КОМПЛЕКСНОГО ПЕРЕМЕННОГО

Рассмотрим на плоскости z кривую C конечной длины, целиком лежащую в некоторой ограниченной области. Разобьем кривую на конечное число частей:

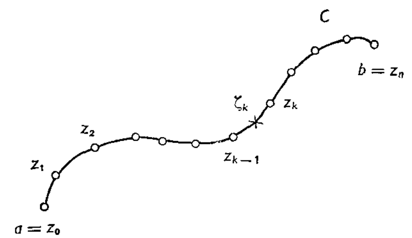
$$a = z_0, z_1, z_2, \ldots, z_n = b.$$

На каждой дуге $\widehat{z_{k-1}, z_k}$ выберем произвольным образом точку ζ_k и составим интегральную сумму

$$I_n = \sum_{k=1}^{n} f(\zeta_k)(z_k - z_{k-1}).$$

Можно доказать, что если функция $f(z)$ непрерывна, то при неограниченном измельчении разбиения кривой C таким образом, чтобы и длина наибольшей из дуг $\widehat{z_{k-1}, z_k}$ стремилась к 0, интегральные суммы I_n независимо от способа разбиения кривой стремятся к определенному пределу. Этот предел называется криволинейным интегралом от функции $f(z)$ по кривой C и обозначается символом

$$\int_C f(z)\,dz$$



Разобьем кривую C в направлении от точки a к точке b на конечное число частей

ПРОСТЕЙШИЕ СВОЙСТВА КРИВОЛИНЕЙНОГО ИНТЕГРАЛА

Перечисленные ниже свойства очевидным образом вытекают из определения интеграла.

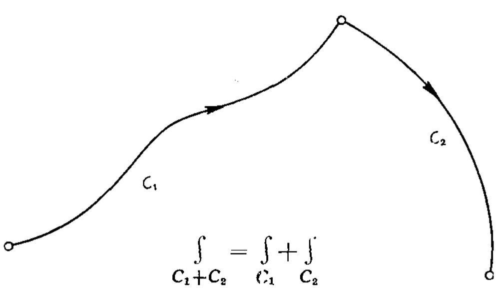
а) Если из конца кривой C_1 выходит кривая C_2, то

$$\int_{C_1+C_2} = \int_{C_1} + \int_{C_2}$$

б) Если интегрирование по кривой C производить в противоположном направлении, то интеграл изменит свой знак на противоположный.

в) $$\int_C kf(z)\,dz = k\int_C f(z)\,dz$$

$$\int_C [f_1(z) + f_2(z)]\,dz = \int_C f_1(z)\,dz + \int_C f_2(z)\,dz$$



$$\int_{C_1+C_2} = \int_{C_1} + \int_{C_2}$$

г) Если $|f(z)| \leqq M$, то

$$\left|\int_C f(z)\,dz\right| \leqq ML,$$

где L — длина кривой C.

СВЕДЕНИЕ К ИНТЕГРАЛУ ОТ ФУНКЦИИ ВЕЩЕСТВЕННОЙ ПЕРЕМЕННОЙ

Предположим, что кривую C можно задать параметрически в виде

$$x = x(t), \qquad y = y(t), \qquad \alpha \leqq t \leqq \beta,$$

где $x(t)$ и $y(t)$ — непрерывно дифференцируемые функции параметра t. Тогда

$$z_k - z_{k-1} = (x_k - x_{k-1}) + i(y_k - y_{k-1}) = \int_{t_{k-1}}^{t_k} [x'(t) + iy'(t)]\,dt = \int_{t_{k-1}}^{t_k} z'(t)\,dt,$$

Отсюда:

$$I_n \sum f(\zeta_k)(z_k - z_{k-1}) = \sum f(z(\tau_k))\,[z(t_k) - z(t_{k-1})] =$$

$$= \sum f(z(\tau_k)) \int_{t_{k-1}}^{t_k} z'(t)\,dt = \int_\alpha^\beta g(t) z'(t)\,dt,$$

где $g(t)$ — функция, принимающая в интервале (t_{k-1}, t_k) постоянное значение $f(z(\tau_k))$ [τ_k — некоторая точка интервала (t_{k-1}, t_k)]. Поскольку функция $f(z(t))$ равномерно непрерывна на отрезке $\alpha \leqq t \leqq \beta$, то при достаточно мелком разбиении функция $g(t)$ сколь угодно мало отличается от функции $f(z(t))$. Следовательно, при неограниченном измельчении разбиения

$$\lim I_n = \lim \int_\alpha^\beta g(t)\,z'(t)\,dt = \int_\alpha^\beta f(z(t))\,z'(t)\,dt,$$

т. е.

$$\boxed{\int_C f(z)\,dz = \int_\beta^\alpha f(z(t))\,z'(t)\,dt.}$$

НЕСКОЛЬКО ПРОСТЫХ ПРИМЕРОВ

1. Пусть $f(z) = c = \text{const}$. Тогда произвольному разбиению соответствует интегральная сумма

$$I_n = \sum_{k=1}^{n} c(z_k - z_{k-1}) = c(z_n - z_0) = c(b - a).$$

Следовательно,

$$\int_C c\,dz = c(b-a).$$

Если кривая C замкнута, то $a = b$ и

$$\int_C C\,dz = 0.$$

2. Пусть $f(z) = z$. Выбирая точки ζ_k сначала в начале, а затем — в конце отрезков разбиения, получим две интегральные суммы:

$$\sum_{k=1}^{n} z_{k-1}\cdot(z_k - z_{k-1}) \to \int_C z\,dz \leftarrow \sum_{k=1}^{n} z_k \cdot (z_k - z_{k-1}).$$

Очевидно, что к тому же пределу стремится и среднее арифметическое рассмотренных интегральных сумм:

$$\frac{1}{2}\sum_{k=1}^{n}(z_{k-1}+z_k)(z_k - z_{k-1}) = \frac{1}{2}\sum_{k=1}^{n}(z_k^2 - z_{k-1}^2) = \frac{1}{2}(z_n^2 - z_0^2).$$

Следовательно,

$$\int_a^b z\,dz = \frac{1}{2}(b^2 - a^2).$$

Если кривая C замкнута, то $\int_C z\,dz = 0$.

3. Проинтегрируем функцию $f(z) = \dfrac{1}{z}$ по окружности $|z| = r$ радиуса r в направлении, противоположном направлению часовой стрелки. Представим окружность в параметрическом виде:

$$x(t) = r\cos t, \qquad y(t) = r\sin t, \qquad 0 \leqq t \leqq 2\pi.$$

Имеем:

$$\int_{|z|=r} \frac{1}{z}\,dz = \int_0^{2\pi} \frac{1}{r(\cos t + i\sin t)}\, r(-\sin t + i\cos t)\,dt =$$

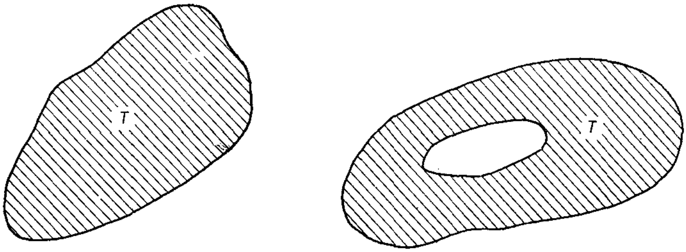
$$= i\int_0^{2\pi} \frac{\cos t - \frac{1}{i}\sin t}{\cos t + i\sin t}\,dt = i\int_0^{2\pi} dt = 2\pi i,$$

так как $\dfrac{1}{i} = -i, \qquad -\dfrac{1}{i} = i.$

ИНТЕГРАЛЬНАЯ ТЕОРЕМА КОШИ

Следующая теорема является одной из важнейших в теории функций комплексного пременного:
если функция $f(z)$ регулярна в некоторой односвязной области T, и C — замкнутая кривая, лежащая в области T, то

$$\int_C f(z)\,dz = 0.$$



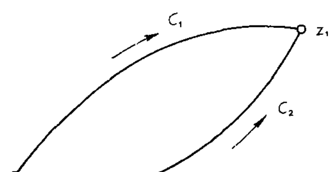
Односвязная область

Область называется односвязной, если для любой простой (то есть не имеющей самопересечений) замкнутой кривой, внутренность этой кривой также принадлежит области.

Приведем несколько следствий из этой теоремы.

1. Рассмотрим кривые C_1 и C_2, принадлежащие области регулярности функции $f(z)$. Если обе кривые выходят из одной точки z_0 и оканчиваются в точке z_1, и если составленная из них замкнутая кривая $C_1 + C_2$ оказывается простой, то

$$\int_{C_1} f(z)\,dz = \int_{C_2} f(z)\,dz.$$



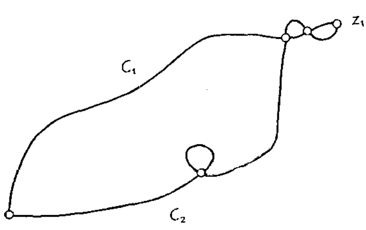
$$\int_{C_1} f(z)\,dz = \int_{C_2} f(z)\,dz$$

Это сразу следует из равенства

$$\int_{C_1} f(z)\,dz - \int_{C_2} f(z)\,dz = \int_{C_1+C_2} f(z)\,dz = 0.$$

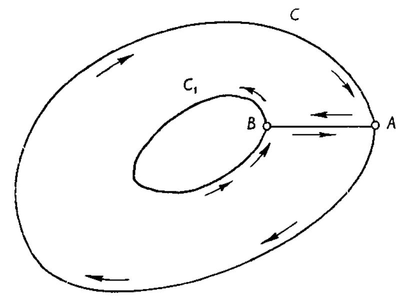
Если кривые C_1 и C_2 имеют самопересечения, а также пересекают друг друга, но кривая $C_1 + C_2$ такова, что может быть разбита на конечное число простых замкнутых кривых, внутренность каждой из которых принадлежит области регулярности функции $f(z)$, то и в этом случае

$$\int_{C_1} f(z)\,dz = \int_{C_2} f(z)\,dz.$$



$$\int_{C_1} f(z)\,dz = \int_{C_2} f(z)\,dz$$

Очень простую и важную формулировку этого следствия можно высказать для случая, когда область регулярности T, которой принадлежит замкнутая кривая $C_1 + C_2$, является односвязной: величина интеграла, взятого по кривой, принадлежащей односвязной области T, в которой интегрируемая функция регулярна, зависит только от начальной и конечной точек, и не зависит от пути интегрирования.



$$\int_C f(z)\,dz = \int_{C_1} f(z)\,dz$$

2. Пусть C и C_1 — простые замкнутые кривые, причем C_1 лежит внутри C, и обе кривые, а также кольцевая область между ними принадлежат области регулярности функции $f(z)$. Тогда

$$\int_C f(z)\,dz = \int_{C_1} f(z)\,dz.$$

В самом деле, если пересечь область между кривыми C и C_1 некоторой линией AB, то получится некоторая замкнутая кривая C', обход которой можно совершить, обходя по разу кривые C и C_1 и дважды пройдя по C'. Применяя теорему Коши к интегралу по кривой C', получим:

$$\int_{C'} = \int_C + \int_A^B - \int_{C_1} + \int_B^A = \int_C - \int_{C_1} = 0,$$

отсюда

$$\int_C = \int_{C_1}.$$

3. Если T — односвязная область и функция $f(z)$ регулярна в области T, то интеграл от функции $f(z)$, взятый по любой кривой, принадлежащей T, зависит только от начальной и конечной точек. Если зафиксировать начальную точку, а конечную точку считать переменной (вместо b обозначим ее через z), то интеграл оказывается функцией конечной точки z:

$$F(z) = \int_a^z f(\zeta)\,d\zeta.$$

Нетрудно доказать, что функция $F(z)$ регулярна в области T и

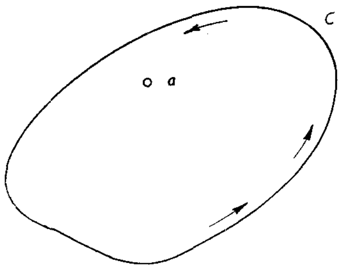
$$F'(z) = f(z).$$

ИНТЕГРАЛЬНАЯ ФОРМУЛА КОШИ

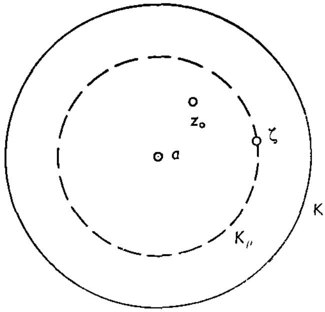
Если C — простая замкнутая кривая, принадлежащая вместе со своей внутренностью области регулярности функции $f(z)$, и точка a лежит внутри C, то

$$f(a) = \frac{1}{2\pi i} \int\limits_C \frac{f(z)}{z-a}\, dz,$$

где обход кривой C совершается против часовой стрелки. Это интегральная формула Коши, являющаяся отправной точкой для большинства серьезных исследований по теории функций комплексного переменного.



$$f(a) = \frac{1}{2\pi i} \int\limits_C \frac{f(z)}{z-a}\, dz$$



На окружности K_r могут быть точки, не принадлежащие к области регулярности функции $f(z)$

Разберем несколько интересных следствий из этой формулы. Рассмотрим окружность K_r радиуса r с центром в произвольной точке a области T регулярности функции $f(z)$, внутренность которой целиком лежит в области T, но на самой окружности уже могут быть точки, не принадлежащие T.

Пусть z_0 — произвольная точка внутри окружности K_r, а $\varrho > 0$ — некоторое число, причем

$$|z_0 - a| < \varrho < r.$$

Тогда окружность K_ϱ радиуса ϱ с центром в точке a целиком лежит внутри окружности K_r. Запишем формулу Коши для интеграла, взятого по окружности K_ϱ:

$$f(z_0) = \frac{1}{2\pi i} \int\limits_{K_\varrho} \frac{f(\zeta)}{\zeta - z_0}\, d\zeta,$$

где ζ — переменная точка, пробегающая по окружности. С другой стороны

$$\frac{1}{\zeta - z_0} = \frac{1}{(\zeta - a) - (z_0 - a)} = \frac{1}{\zeta - a} \cdot \frac{1}{1 - \frac{z_0 - a}{\zeta - a}}.$$

Поскольку абсолютная величина дроби $\frac{z_0 - a}{\zeta - a}$ меньше 1, то последняя дробь в записанном выше равенстве является суммой бесконечного геометрического ряда

$$1 + \frac{z_0 - a}{\zeta - a} + \left(\frac{z_0 - a}{\zeta - a}\right)^2 + \ldots + \left(\frac{z_0 - a}{\zeta - a}\right)^n + \ldots = \sum_{n=0}^{\infty} \left(\frac{z_0 - a}{\zeta - a}\right)^n.$$

Следовательно,

$$\frac{1}{\zeta - z_0} = \sum_{n=0}^{\infty} \frac{(z_0 - a)^n}{(\zeta - a)^{n+1}}.$$

Этот ряд равномерно сходится на окружности K_ϱ а значит, его можно интегрировать почленно. Таким образом из формулы Коши получаем:

$$f(z_0) = \frac{1}{2\pi i} \sum_{n=0}^{\infty} \left(\int_{K_\varrho} \frac{f(\zeta)}{(\zeta - a)^{n+1}} \, d\zeta \right) \cdot (z_0 - a)^n.$$

Если ввести обозначение

$$c_n = \frac{1}{2\pi i} \int_{K_\varrho} \frac{f(\zeta)}{(\zeta - a)^{n+1}} \, d\zeta,$$

то

$$f(z_0) = \sum_{n=0}^{\infty} c_n (z_0 - a)^n.$$

Коэффициенты этого ряда, казалось бы, зависят от выбора точки z_0, так как z_0 присутствует в условии, наложенном на радиус ϱ. Однако функции

$$\frac{f(z)}{(z - a)^{n+1}}$$

регулярны в области, которая получается, если из области регулярности функции $f(z)$ исключить точку a. Поэтому интегралы от этих функций, взятые по окружности K_ϱ, соответственно равны интегралам, взятым по любой простой замкнутой кривой C, принадлежащей области регулярности и не проходящей через точку a. (См. следствие 2 из интегральной теоремы Коши). Отсюда следует, что коэффициенты c_n нашего ряда не зависят от z_0. Поскольку точка z_0 выбиралась произвольно, то для любой точки внутри окружности имеет место разложение

$$f(z) = \sum_{n=0}^{\infty} c_n(z-a)^n,$$

где коэффициенты даются формулой

$$c_n = \frac{1}{2\pi i} \int_C \frac{f(\zeta)}{(\zeta - a)^{n+1}}\, d\zeta\,.$$

Здесь C — простая замкнутая кривая, обход которой совершается против часовой стрелки, причем точка a лежит внутри C.

Подведем итог: если функция $f(z)$ регулярна в области T, и a — произвольная точка T, то, рассматривая окружность K_r радиуса r с центром в точке a, внутренность которой принадлежит области T, функцию $f(z)$ можно разложить в степенной ряд

$$f(z) = \sum_{n=0}^{\infty} c_n(z-a)^n\,; \qquad c_n = \frac{1}{2\pi i} \int_C \frac{f(\zeta)}{(\zeta - a)^{n+1}}\, d\zeta\,,$$

который сходится внутри окружности K_r.

Поскольку сумма сходящегося степенного ряда — бесконечно дифференцируемая функция, то коэффициенты c_n могут быть получены из функции $f(z)$ как коэффициенты ряда Тейлора:

$$c_n = \frac{1}{n!} f^{(n)}(a) \qquad (n = 0, 1, 2, \ldots)$$

Сравнивая формулы

$$f(z) = \frac{1}{2\pi i} \sum_{n=0}^{\infty} \left(\int_C \frac{f(\zeta)}{(\zeta - a)^{n+1}}\, d\zeta \right) (z-a)^n$$

и

$$f(z) = \sum_{n=0}^{\infty} c_n(z-a)^n, \qquad c_n = \frac{f^{(n)}(a)}{n!}\,,$$

получаем формулы для производных функции $f(z)$:

$$f^{(n)}(a) = \frac{n!}{2\pi i} \int_C \frac{f(\zeta)}{(\zeta - a)^{n+1}}\, d\zeta \qquad (n = 0, 1, 2, \ldots)$$

Из разложения в степенной ряд, получаемого с помощью формулы Коши, следует, в частности, что функция, дифференцируемая в некоторой области, бесконечно дифференцируема в этой области. Для функций вещественного переменного это неверно. Здесь нетрудно указать пример такой функции, которая имеет первую, но не имеет второй производной. Кроме того, в вещественном случае из дифференцируемости не следует, что функция может быть разложена в степенной ряд. Так, например, функция

$$y = \begin{cases} e^{-\frac{1}{x^2}}, & \text{если} \quad x \neq 0 \\ 0 & \phantom{\text{если}} \quad x = 0 \end{cases}$$

бесконечно дифференцируема, но в окрестности 0 не разлагается в степенной ряд, так как все ее производные там обращаются в нуль.

КРАТКИЙ ОБЗОР ИСТОРИИ МАТЕМАТИЧЕСКОГО АНАЛИЗА

Уже древнегреческие математики использовали для решения некоторых геометрических задач методы, похожие на методы математического анализа. Так, например, Архимед (287 — 212 гг. до н. э.) при вычислении площадей некоторых фигур и при определении объема шара использовал по существу интегральное исчисление, хотя, естественно, не знал его общих методов. Тем не менее ни Архимед, ни другие древнегреческие математики не осознали значения этих методов, и даже не увидели возможности использовать их для решения общих задач на определение площадей и объемов.

Развитие математического анализа в строго научном смысле было вызвано развитием производительных сил во времена зарождения капитализма. Великие географические открытия, например, привели к развитию мореплавания, вследствие этого было изобретено множество новых машин, исследование которых повлекло за собой отыскание новых путей и методов в теоретической механике, в результате возникла необходимость в новом математическом аппарате.

Одним из основателей математического анализа стал английский ученый И. Ньютон (1643 — 1727). Исследования в области механики привели его к проблемам дифференциального и интегрального исчисления. Не случайно (история знает немало подобных примеров), что одновременно с Ньютоном проблемами анализа занимался и немецкий математик Г. В. Лейбниц (1646 — 1716). Поскольку дифференциальное и интегральное исчисление имеет дело с бесконечно малыми величинами, то этот раздел математики иногда называют также анализом бесконечно малых.

К концу XVII века дальнейшее развитие средств производства, а также познание сути природных явлений и установление общих законов, которым они подчиняются, стало насущной необходимостью, и только анализ бесконечно малых мог дать единственно приемлемый метод для их точного описания. Все это предопределило открытие, сделанное двумя великими европейскими учеными.

XVIII век ознаменовался дальнейшим развитием математики, что было вызвано бурным техническим прогрессом. Работы целого ряда выдающихся математиков значительно провинули вперед теорию и, в частности, математический анализ. Из этого периода необходимо отметить швейцарского математика Л. Эйлера (1707—1783), который свыше 30 лет проработал в России, являлся членом Петербургской Академии наук и систематизировал анализ бесконечно малых. Систематизацией уже имевшихся результатов, а также дальнейшим развитием теории занимались и многие французские математики. Среди них следует упомянуть известного энциклопедиста Ж. Даламбера (1717 — 1783), а также Л. Лагранжа (1736 — — 1818), П. Лапласа (1749 — 1827), А. Лежандра (1752 — 1833) и Ж. Фурье (1768 — 1830), который добился замечательных результатов в теории тригонометрических рядов.

В XIX веке развитие математического анализа продолжилось, хотя математики считали своей основной задачей разрешить противоречия, возникшие в ту пору. Дело в том, что некоторые математики в различных областях анализа пришли к ошибочным, логически противоречивым результатам. Очевидные противоречия послужили причиной горячих споров, что привело затем к критическому пересмотру методов и четкому логическому построению математического анализа. В частности, были уточнены понятие предельного перехода и основанные на нем понятия производной и интеграла, независимых от «бесконечно малых» величин. Здесь следует упомянуть немецкого математика К. Гаусса (1777 — 1855), французского математика О. Коши (1780 — 1856), который стал основателем теории функций комплексного переменного и определил понятие предела; норвежского математика Н. Абеля (1802 — 1829); немецких математиков: П. Дирихле (1805 — 1859), К. Вейерштрасса (1815 — 1897), Р. Дедекинда (1831 — 1916) и Г. Кантора (1845 — 1918); русского математика П. Л. Чебышева (1821 — 1866), создателя конструктивной теории функций; немецкого математика Б. Римана (1826 — 1866); русского математика С. В. Ковалевскую (1850 — 1891); английских математиков Г. Харди (1877 — — 1947) и Д. И. Литлвуда (1885 —); немецкого математика Д. Гильберта (1862 — 1943). Деятельность последних трех проходила большей частью уже в XX веке. Французский математик А. Лебег (1875 — 1941), обобщивший понятие интеграла, также начал свою деятельность в конце XIX века.

Важнейшие труды, касающиеся критики методов и стройного логического построения математического анализа, принадлежат Дирихле, Вейерштрассу, Дедекинду и Кантору.

В нашем столетии важных результатов в математическом анализе добился французский математик Э. Борель (1871 — 1956).

Очень интенсивно проходят исследования в области математического анализа в Советском Союзе. Советские математики достигли значительных результатов в этой области, среди них немало ученых с мировым именем. Наиболее известные из них: Н. К. Бари, С. Н. Бернштейн, М. В. Келдыш, А. Н. Колмогоров, А. М. Ляпунов, А. А. Марков, Е. Г. Петровский, и многие другие.

В таком коротком обзоре, естественно, невозможно перечислить даже имен всех тех замечательных ученых, которые внесли свой вклад в развитие математического анализа. Следует, однако, упомянуть имена некоторых из венгерских математиков, которые своей деятельностью заграницей или в Венгрии заслужили мировое признание. Й. Нейман, Дёрдь

Пойа, Тибор Радо, Отто Сас, Габор Сегё — в США, Михай Фекете — в Израиле, а Марцель Рисс — в Швеции. Все они работали или продолжают работать и относятся к числу виднейших специалистов в области математического анализа.

В Венгрии исследования в области математического анализа достигли высокого уровня. В этом большая заслуга Дьюлы Кёнига (1849 — 1913). Особо следует упомянуть также имена Липота Фейера (1880 — 1959) и Фридьеша Рисса (1880 — 1956), которые своими работами в области математического анализа приобрели мировую известность.

Немало венгерских математиков занимается анализом и в наши дни. Среди них Дёрдь Алексич (1899 —), Акош Часар (1924 —), Бела Сёкефалви Надь (1913 —), Пал Туран (1911 —) и другие. Замечательные результаты достигнуты в этой области и молодым поколением венгерских математиков.

ТЕОРИЯ МНОЖЕСТВ

АЛГЕБРА МНОЖЕСТВ

ПОНЯТИЕ МНОЖЕСТВА

Как в повседневной, так и в научной жизни часто говорят о чертах какого-либо коллектива, совокупности некоторых вещей. Так, например, можно говорить о зрителях театра, кино, о совокупности венгерских граждан, о коллективе социалистических стран, о совокупности естественных чисел, о совокупности точек внутри некоторого круга и т. д.

Понятие *множества* в математике выведено из понятия совокупностей, образуемых из ограниченного количества предметов, сведенных к единице. Предметы, собранные в множество, называются *элементами* множества. Понятия «множество» и «элемент» считаются основными понятиями (как в геометрии понятия точки, прямой и плоскости) и не сведены к другим (математическим или логическим) понятиям путем применения формального определения. Приведенными ниже замечаниями эти два понятия и их взаимные связи поясняются достаточно точно для того, чтобы их можно было однозначно применять.

Раздел математики, занимающийся множествами, называется *теорией множеств*. Ее основоположником был немецкий математик Георг Кантор (1845—1918).

По оригинальной, так называемой *наивной* точке зрения, элементами множества могут быть любые предметы. Каждое множество считается самостоятельной, осмысленной вещью, как бы осмысленной оболочкой его элементов. Множество считается известным, если заданы его элементы; множество определяется раз и навсегда заданием его элементов; множества не зависят от времени. Все это кратко суммируется в определении, являющемся аксиомой, основным законом:

Множество однозначно определяется его элементами.

При доказательстве теорем относительно множеств будем ссылаться на эту аксиому и законы логики.

Для обозначения множеств обычно применяются заглавные латинские буквы. Выражение

$$d \in H$$

обозначает, что предмет d является элементом H (читается: «d является элементом H»), а отрицание, то есть предмет d *не является* элементом множества H обозначается выражением

$$d \notin H$$

(читается: «d не является элементом H»).

Тождество множеств. Если два множества тождественны (равны), то на основе понятия тождественности элементы обоих множеств одинаковы.

Если о двух множествах нам известно, что их элементы тождественны, то по нашей аксиоме эти два множества тождественны, так как множество однозначно определяется его элементами. Поэтому:

два множества тождественны тогда и только тогда, если их элементы одинаковы.

В определениях, касающихся *геометрических мест*, всегда присутствует отождествление множеств, заданных двумя разными определениями.

Пример. Перпендикулярная линия, пересекающая в середине отрезок прямой, является геометрическим местом точек, расположенных на одинаковом расстоянии от двух концов отрезка. Это означает следующее: в плоскости множество точек перпендикулярной линии, пересекающей в середине отрезок прямой, тождественно множеству точек, расположенных на одинаковом расстоянии от обоих концов отрезка.

Множество часто задается в следующем виде: элементы множества заключаются внутри фигурных скобок ($\{\ldots\}$). Подобной «записью» может быть конкретное перечисление элементов множества (запись их названия или обозначения) или задание такого определения, которым элементы множества однозначно задаются.

Пример задания путем определения:

{гласные звуки слов «ТЕОРИЯ МНОЖЕСТВ»}

или {гласные звуки слова «ПОЗНАЕМ»}.

Оба множества могут быть заданы также путем перечисления их элементов:

$$\{Е,\ О,\ И,\ Я,\ О,\ Е\}, \qquad \{О,\ А,\ Ё\}.$$

Можно видеть, что элементы «*Е*» и «*О*» в левом множестве перечислены дважды. Но в то же время один гласный звук «*Е*» тождествен другому гласному звуку «*Е*» (то же самое в отношении гласного звука «*О*»); многократное повторение одного и того же предмета в теории множеств не подразумевает образование нового предмета.

Вывод из примера:

один предмет в одном множестве является элементом только один раз (даже если предмет повторяется несколько раз).

Тождественные (равные) множества связываются знаком равенства («=»):

$$\{A,\ A,\ E,\ E\} = \{A,\ E\}.$$

Если один из тождественных предметов является элементом множества, то другой из тождественных предметов также будет элементом данного множества т. е. если $a = b$ и $a \in H$, то и $b \in H$. Это вытекает из понятия тождественности. Например, если поэт, написавший стихотворение «Больше воздуха!», является элементом множества моих любимых поэтов, то Аттила Йожеф также является элементом множества моих любимых поэтов, так как

поэт, написавший стихотворение
«Больше воздуха!» = Аттила Йожеф.

Не существует никакого ограничения в отношении того, насколько «много» (или «мало») элементов может быть в одном множестве: в одном множестве может быть любое, даже *бесконечное* количество элементов. Множество считается конечным, если имеет конечное количество элементов, т. е., если число его элементов может быть задано естественным числом.

Множество считается бесконечным, если имеет бесконечное количество элементов, т. е. если число его элементов не может быть задано никаким естественным числом. (Пример бесконечного множества: множество естественных чисел, множество точек отрезка прямой и т. д.)

В множестве может быть также и любое «малое» количество элементов. Таким образом, существуют, например, множества с одним элементом.

Пустое множество. Определением

$$\{\text{действительные корни уравнения } x^2 + 1 = 0\}$$

задается такое множество, в котором вообще нет ни одного элемента (см. главу «Алгебра»). Существование такого множества не противоречит нашему основному закону, так как элементы этого множества определены (по возможности) самым совершенным образом: ни один предмет не является его элементом. Такие множества называются *пустыми множествами*. Множество городов с населением выше 3 миллионов в Венгрии также представляет собой пустое множество.

В то же время все пустые множества тождественны: то, что является элементом одного множества, одновременно является и элементом другого множества, так как ни одно пустое множество не имеет элемента. Таким образом, существует всего одно пустое множество. Суммируем:

пустым называется множество, у которого ни один предмет не является элементом.

Пустое множество обозначается символом «$\emptyset$» (читается «пустое»).

Элементами множества могут быть любые вещи, в том числе и множества:

$$\{\emptyset\}, \qquad \{\{\emptyset\}\}, \qquad \{\{\{\emptyset\}\}\}, \qquad \ldots$$

Все эти множества являются множествами с одним элементом. Единственным элементом первого множества является пустое множество, единственным элементом второго — первое множество (т. е. множество, у которого единственный элемент — пустое множество), в третьем множестве единственным элементом является второе множество (т. е. множество, единственный элемент которого такое множество, у которого единственный элемент — пустое множество). Начатая последовательность изображается таким образом, что последнее полученное множество заключается в скобках. — В следующих примерах представлены такие множества с несколькими элементами, в которых элементы также являются множествами:

$$\{\emptyset, \{\emptyset\}\}, \qquad \{\emptyset, \{\emptyset\}, \{\{\emptyset\}\}\}, \qquad \{\emptyset, \{\emptyset\}, \{\emptyset, \{\emptyset\}\}\}.$$

ПОДМНОЖЕСТВО, ДЕЙСТВИТЕЛЬНАЯ ЧАСТЬ

Множество A считается подмножеством множества B, если каждый элемент A является и элементом B, что обозначается выражением $A \subseteq B$.

Примеры:

Множество живых организмов на Земле является подмножеством живых организмов солнечной системы: любое живое существо на Земле одновременно является живым организмом нашей солнечной системы. В настоящее время неизвестно, существуют ли такие живые организмы в солнечной системе, которых нет на земном шаре.

Множество тех парных чисел, которые могут быть приведены как сумма двух неделимых чисел, наверняка является подмножеством множества парных чисел больше двух. Парное число, больше двух, которое не может быть приведено как сумма двух неделимых чисел, неизвестно, но до сих пор не удалось доказать невозможности существования такого парного числа.

Понятие «части» (подмножества) в теории множеств удобно для обозначения таких соотношений. Настоящее понятие отличается от обычного понятия «части»; по последнему понятию часть всегда «меньше», чем целое. Но по понятию «части» в теории множеств «целое» также входит в понятие части. Точнее говоря:
каждое множество является элементом самого себя, так как каждый элемент A является элементом A, значит, $A \subseteq A$. (Рефлексивность «$\subseteq$»). Далее: *пустое множество является частью каждого множества*, так как каким бы ни было множество A, каждый элемент пустого множества является элементом A (так как пустое множество вообще не содержит элементов).

Обычному понятию «части» соответствует *действительная часть*.

Считается, что
множество A является действительным подмножеством (действительной частью) множества B, если A является частью B, но не тождественно с ним, что обозначается символом $A \subset B$. (Короче говоря, $A \subset B$, если $A \subseteq B$ и $A \neq B$; символом «$\neq$» обозначается «не равно»).

Примеры:

Множество неделимых чисел является *действительной частью* множества естественных чисел.

Множество параллелограммов является *действительной частью* множества четырехугольников.

Множество дифференцируемых на данном интервале функций является *действительной частью* множества интегрируемых на этом интервале функций.

Другие эквивалентные определения действительной части:

$A \subset B$, если $A \subseteq B$, но $B \nsubseteq A$. Символом «$\not\subset$» обозначается «не является частью».

$A \subset B$, если $A \subseteq B$ и существует такой предмет d, что $d \notin A$, но $\partial \in B$.

Нетрудно понять, что эти определения равноценны предыдущему. Так же легко показать правильность следующих соотношений:

Если $A \subseteq B$ и $B \subseteq A$, то $A = B$. (Те предметы, которые являются частью друг друга, тождественны между собой.)

Пример: Пусть множество корней уравнения E обозначается через $R(E)$. Как известно, уравнение E_1 является следствием уравнения E_2, если каждый корень уравнения E_2 является корнем уравнения E_1, т. е. если $R(E_2) \subseteq R(E_1)$. Если уравнения E_1 и E_2 являются следствием друг друга, то E_1 и E_2 являются *эквивалентными* уравнениями. Если эквивалентность уравнений E_1 и E_2 выразить с помощью множества корней, то $R(E_1) \subseteq R(E_2)$ и $R(E_2) \subseteq R(E_1)$, то есть, $R(E_1) = R(E_2)$.

Если $A \subset B$, то $A \subseteq B$. (Действительная часть сама всегда является частью.)

$A \subset A$ не является правильным утверждением. *(Иррефлексивность «$\subset$».)*

Если $A \subset B$, то $B \not\subset A$. *(Асимметрия «$\subset$».)*

Если $A \subset B$ и $B \subset C$, то $A \subset C$. *(Транзитивность «$\subset$».* Это утверждение действительно и в том случае, если применять знак $\subseteq$ вместо знака $\subset$.)

ОПЕРАЦИИ МНОЖЕСТВ

Союз множеств. Под понятием союза (или соединения) двух множеств A и B подразумевается множество всех тех предметов, которые являются элементами по крайней мере одного из множеств A и B; это множество обозначается символом $A \cup B$ (читается «A союз B»). A и B являются членами союза. Следовательно, элементами $A \cup B$ являются те и только те предметы, которые являются элементами A, B или обоих.

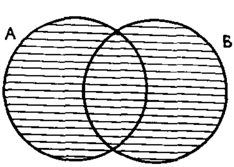
Пример. Пусть $f(x)$ — любая действительная функция и пусть через A, B, C, обозначаются области существования функций $f(x)$, $\sqrt{f(x)}$, $\sqrt{-f(x)}$ соответственно. Ясно, что

$x \in B$ тогда и только тогда, если $f(x) \geqq 0$ и

$x \in C$ тогда и только тогда, если $f(x) \leqq 0$,

и поэтому:

$$B \cup C = A\,.$$

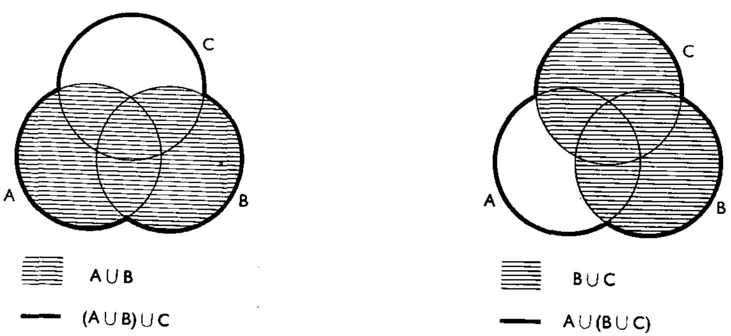


$A \cup B$

Легко увидеть, что если A и B являются ограниченными множествами без общих элементов, то количество элементов $A \cup B$ = (количество элементов A) + (количество элементов B). На основе этих соотношений операция союза часто называется суммированием множеств. Операции множеств и связанные с ними соотношения представляются наглядно с помощью диаграмм Венна, названных именем английского математика Дж. Венна (1834—1923). На этих диаграммах любые множества изображаются кругами, пересекающими друг друга, исходя из того, что внутренними точками круга изображаются элементы множества. Общей частью двух кругов, пересекающих друг друга, представляются возможные общие элементы двух множеств. На рисунке представлена диаграмма операции союза; заштрихованной площадью иллюстрируется операция $A \cup B$ (стр. 540).

Легко увидеть (на основе определений и диаграмм) следующие тождества операции союза. То, что этими равенствами выражаются тождества, означает, что всегда получаются правильные равенства, если поставить любые множества на место букв A, B, C в тождествах.

$$A \cup B = B \cup A \quad (\textit{коммутативность}),$$
$$A \cup A = A \quad (\textit{идемпотенция}).$$
$$A \cup \emptyset = A$$
$$(A \cup B) \cup C = A \cup (B \cup C) \quad (\textit{ассоциативность}),$$

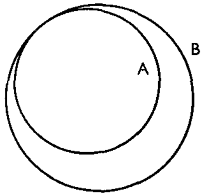


Операция союза является коммутативной и ассоциативной (как в арифметике операции суммирования и перемножения). На основе этих соотношений союз с любым количеством членов пишется без скобок независимо от того, в какой последовательности проводится соединение множеств, присутствующих в союзе.

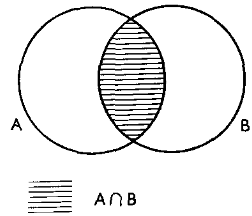
Также действительны и нижеприведенные соотношения:

$A \subseteq A \cup B, \qquad B \subseteq A \cup B.$

$A \subseteq B$ тогда и только тогда, если $A \cup B = B$ (см. рисунок).



B = A ∪ B

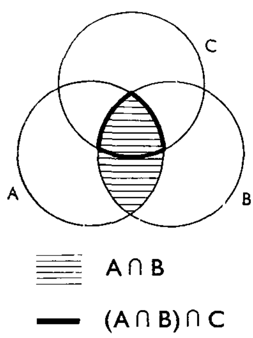


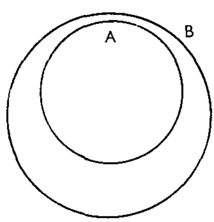
Под термином разрез (или общая часть) множеств A, B подразумевается множество всех тех предметов, которые одновременно являются элементами и множества A, B. Такое множество обозначается символом $A \cap B$ (читается: «A разрез B»). A и B являются членами разреза. (Разрез иногда называется произведением.)

Примеры: *Разрезом* множества ромбов и множества прямоугольников является множество квадратов. Область существования суммы (или произведения) двух функций является разрезом областей существования этих двух функций.

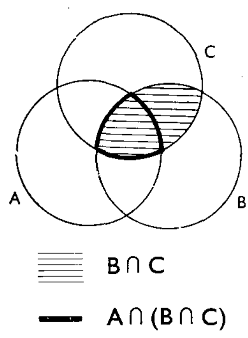
Диаграмма операции разреза представлена на рисунке. Основные тождества этой операции представлены ниже.

$A \cap B = B \cap A$	*(коммутативность)*,
$A \cap A = A$	*(идемпотенция)*,
$A \cap \emptyset = \emptyset$	
$(A \cap B) \cap C = A \cap (B \cap C)$	*(ассоциативность;* см. рисунок).





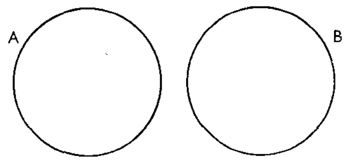
A ∩ B = A



Действительны также следующие соотношения:

$$A \cap B \subseteq A,$$
$$A \cap B \subseteq B.$$
$$A \subseteq B,$$
тогда и только тогда, если $A \cap B = A$ (см. рисунок).

Если $A \cap B = \emptyset$, т. е. если A и B не имеют общих элементов, то считается, что A и B являются *посторонними* (или *чужими)* множествами (см. рисунок).

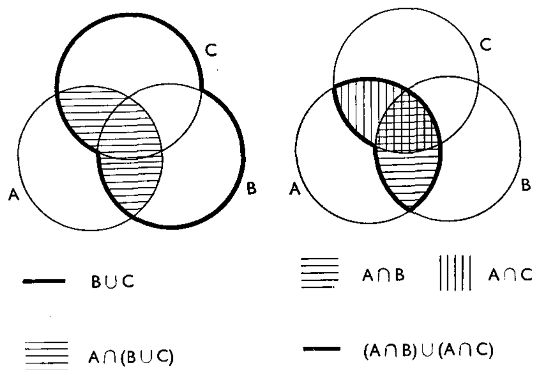


$A \cap B = \emptyset$

Как видно из вышеуказанных тождеств, операция разреза также является коммутативной и ассоциативной (подобно операции союза). Разрез относительно союза, а союз относительно разреза является *дистрибутивным* (как в арифметике умножение является дистрибутивным относительно суммирования):

$$A \cap (B \cup C) =$$
$$= (A \cap B) \cup (A \cap C),$$

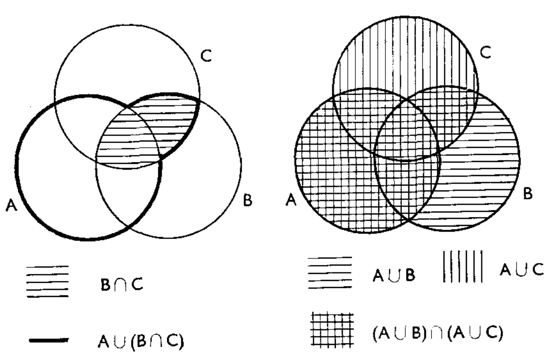
$$A \cup (B \cap C) =$$
$$= (A \cup B) \cap (A \cup C).$$

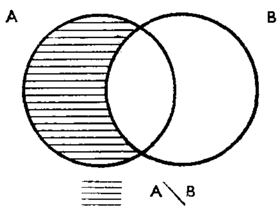


(См. рисунки.) Обобщения этих тождеств приведены аналогично правилам умножения относительно многочленов в арифметике.

Приводятся еще два интересных тождества (так называемые правила поглощения):

$$A \cap (A \cup B) = A;$$
$$A \cup (A \cap B) = A.$$





Эти соотношения получаются на основании первого рисунка (стр. 540) и второго рисунка (стр. 542).

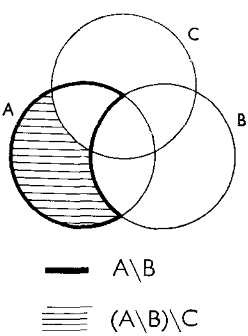
Разность множеств. Под определением разности предстоящего члена A и последующего члена B подразумевается множество тех элементов множества A, которые *не* являются элементами множества B, что обозначается символом $A\setminus B$ (читается: «A минус B»).

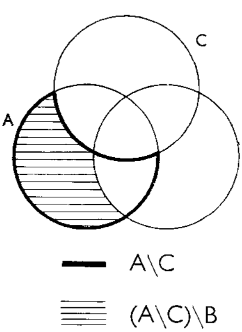
Диаграмма этой операции представлена на рисунке.

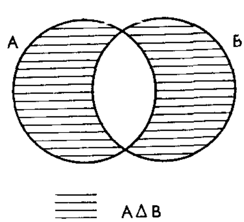
Пример: Если областью существования функции $f(x)$ является A, а множеством ее корней — B, то область существования функции $\dfrac{1}{f(x)}$ — множество $A\setminus B$.

Ниже приведены основные соотношения операции разности:

$$A\setminus\emptyset = A \qquad\qquad \emptyset\setminus A = \emptyset$$
$$A\setminus A = \emptyset$$
$$(A\setminus B)\cap(B\setminus A) = \emptyset$$
$$(A\setminus B)\cup B = A\cup B$$
$$(A\setminus B)\cap A = A\setminus C \qquad\qquad (A\setminus B)\cap B = \emptyset$$
$$(A\setminus B)\setminus C \doteq (A\setminus B)\setminus B \qquad\qquad \text{(См. рисунок)}$$
$$A\setminus B \subseteq A.$$



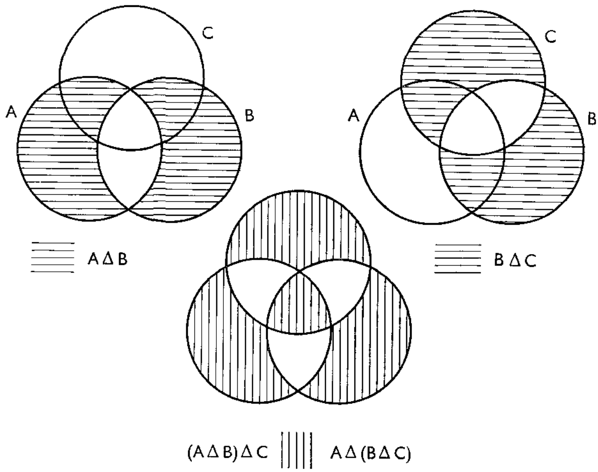




На основе определений нетрудно убедиться в правильности вышеуказанных соотношений; большую их часть можно видеть на предыдущем рисунке.

Симметричная разность. Под симметричной разностью двух множеств A и B подразумевается множество $(A \setminus B) \cup (B \setminus A)$; оно обозначается символом $A \triangle B$ (читается: «A дельта B»).

A и B являются членами симметричной разности. Диаграмма этой операции представлена на рисунке.



Ниже приведены несколько соотношений этой операции:

$$A \triangle B = (A \cup B) \setminus (A \cap B)$$
$$A \triangle B = B \triangle A$$
$$(A \triangle B) \triangle C = A \triangle (B \triangle C)$$
(См. рисунок.)
$$A \triangle A = \emptyset \qquad A \triangle \emptyset = A$$
$$A \triangle (A \triangle B) = B$$
$$A \triangle B \subseteq A \cup B$$

Упорядоченная пара. Два любых предмета a и b представляют собой упорядоченную пару, если предварительно задано, какой из них считается первым и какой — вторым. Символом «(a, b)» обозначается такая упорядоченная пара, первым членом которой является a, а вторым — b (a и b могут быть тождественны). — Например, дроби могут рассматриваться как упорядоченные пары, образуемые из естественных чисел. Первым членом такой пары считается числитель, вторым — знаменатель. Таким образом, $\frac{2}{3}$ соответствует упорядоченной паре (2,3), а $\frac{3}{2}$ — (3,2), или $\frac{4}{4}$ — (4,4). То же относится и к понятию упорядоченных трех, четырех, . . ., n - чисел ($n \geqq 2$).

Непосредственное произведение. Под определением непосредственного произведения двух множеств A и B подразумевается множество всех тех упорядоченных пар, в которых первый член является элементом A, а второй член — элементом B (обозначается символом $A \times B$).

Пример:

Если $A = \{a, e, i\}$, $B = \{b, c\}$, то

$$A \times B = \{(a, b), (a, c), (e, b), (e, c), (i, b), (i, c)\} .$$

Далее:

$$B \times B = \{(b, b), (b, c), (c, b), (c, c)\} .$$

В координатной геометрии точки плоскости характеризуются парами чисел, а точки пространства — тремя числами. Это означает, что если A обозначает множество точек числовой оси, то $A \times A$ — множество точек плоскости, а $(A \times A) \times A$ — множество точек пространства.

Непосредственное произведение обычно не является коммутативным (так как первый член $A \times B$ происходит из множества A, а $B \times A$ — из множества B) и также не является ассоциативным. В то же время непосредственное произведение является дистрибутивным относительно операции союза:

$$A \times (B \cup C) = (A \times B) \cup (A \times C) ; \quad (A \cup B) \times C = (A \times C) \cup (B \times C) .$$

$A \times \emptyset = \emptyset = \emptyset \times A$, так как пустое множество не имеет элементов и поэтому его элементы не могут быть членами упорядоченных пар. — Таким образом, если A и B являются *ограниченными* множествами, то

количество элементов $A \times A$ = (количество элементов A) · (количество элементов B).

АЛГЕБРА СТЕПЕННЫХ МНОЖЕСТВ

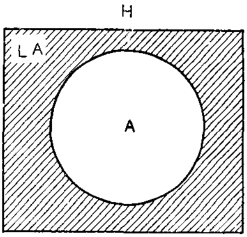
Можно доказать (путем применения полной индукции), что если ограниченное множество имеет n элементов, то количество всех его подмножеств составляет 2^n (подмножествам принадлежат и пустое множество, и данное). Например, количество элементов пустого множества равно 0, количество его частей — 1 (оно имеет единственную часть: самое себя) и $2^0 = 1$. Множество $\{a\}$ с одним элементом имеет две части: $\emptyset$ и самое себя, и $2^1 = 2$. Частями множества $\{a, b\}$ с двумя элементами являются:

$$\emptyset, \qquad \{a\}, \qquad \{b\}, \qquad \{a, b\} ,$$

количество частей — 4 и $2^2 = 4$ и т. д.

Множество всех частей любого (ограниченного или бесконечного) множества H называется степенным множеством множества H и обозначается через 2^H.

Комплемент. Пусть H — зафиксированное непустое множество. На подмножествах множества H интерпретируется новая операция: операция образования комплемента. Если $A \subseteq H$, то под определением *комплемента* A подразумевается множество $H \setminus A$, что обозначается символом $\complement A$. Ясно, что $\complement A \subseteq H$.



На рисунке представлена диаграмма комплемента: множество H изображается четырехугольником, а части множества H — кругами внутри четырехугольника.

Пример: Пусть множество H — область существования действительной функции $f(x)$, а множество A — область существования функции $\sqrt{f(x)}$. Очевидно, что $A \subseteq H$. В данном случае областью существования функции $\sqrt{-\frac{1}{f(x)}}$ является именно множество $\complement A$.

Можно показать правильность следующих тождеств:

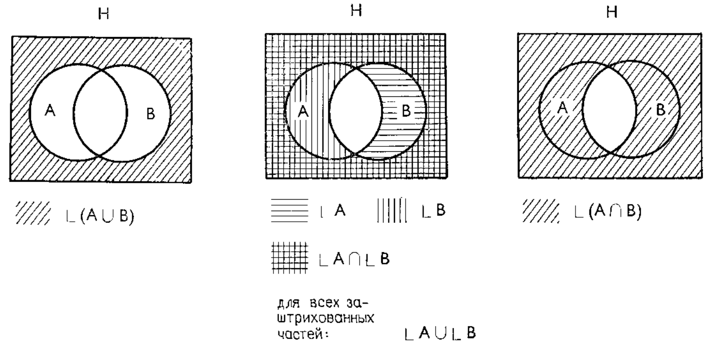
$$\complement H = \emptyset, \qquad \complement \emptyset = H,$$
$$\complement\complement A = A.$$

Если $A \subseteq H$ и $B \subseteq H$, то $A \cup B \subseteq H$ и $A \cap B \subseteq H$. Таким образом, если применять операции союза, разреза и образования комплемента относительно подмножеств множества H, то всегда получается одно подмножество множества H. Этот факт выражается таким образом, что множество частей множества H (т. е. степенное множество H) образует *замкнутую систему* относительно операций $\cup, \cap, \complement$. Множество 2^H, вместе с операциями $\cup, \cap, \complement$, определенными относительно элементов (т. е. относительно частей H) и вместе с действительными тождествами, называется *алгеброй степенных множеств H* (или *H-алгеброй*), при условии, что H не представляет собой пустое множество.

Естественно, что тождества, относящиеся к операциям союза и разреза, действительны и в H-алгебре.

Ниже показаны тождества, содержащие смешанно три операции:

$A \cup \complement A = H$	$A \cap \complement A = \emptyset$
$A \cup (B \cap \complement B) = A$	$A \cap (B \cup \complement B) = A$
$\complement (A \cup B) = \complement A \cap \complement B$	$\complement (A \cap B) = \complement A \cup \complement B$



Последние два тождества называются тождествами де Моргана (по имени английского математика А. де Моргана (1806—1878)). Эти тождества наглядно представлены на рисунке.

Знаки H и $\emptyset$ играют в тождествах определенную, постоянную роль и называются *константами*. Буквами A, B, C обозначается любая часть множества H; эти буквы называются *переменными*. Естественно, что для обозначения переменных могут быть применены любые буквы, за исключением знаков *констант*. *Выражениями* в H-алгебре считаются константы H и $\emptyset$, переменные и все те ограниченные последовательности знаков, которые выводятся из предыдущих путем регулярного применения (ограниченный раз) операций $\cup$, $\cap$, $\llcorner$. Во всех вышеуказанных тождествах по обеим сторонам знака «$=$» стоят выражения.

Замена. Так как переменные в тождествах представляют собой любые части степенного множества H, то из одного тождества получится другое, если в нем любая переменная (во всех местах тождества) *заменяется* любым выражением H-алгебры. *(Правило замены.)* Например, из тождества $\llcorner\llcorner A = A$ путем замены A на выражение $B \cup C$ получится тождество $\llcorner\llcorner(B \cup C) = B \cup C$.

Дополнение. Из одного тождества получится другое тождество и в том случае, если такое выражение как связная часть, которое присутствует на одной стороне тождества, *дополняется* (заменяется) тождественным с ним выражением. *(Закон дополнения.)* Например, в тождестве $\llcorner(A \cup B) = \llcorner A \cap \llcorner B$ путем применения тождества $A \cup B = B \cup A$, присутствующее на левой стороне уравнение как часть уравнения $A \cup B$, дополняется выражением $B \cup A$, и, таким образом, получится тождество $\llcorner(B \cup A) = \llcorner A \cap \llcorner B$. (Если дополняемое выражение повторяется в тождестве несколько раз, то при замене не следует повторять дополнение.)

Тождества H-алгебры. Можно доказать, что любое тождество H-алгебры может быть приведено на основе следующих тождеств путем ограниченно многократного применения замены и дополнения.

$$
\boxed{
\begin{array}{ll}
A \cup A = A & A \cup B = B \cup A \\
(A \cup B) \cup C = A \cup (B \cup C) & A \cup (B \cap C) = (A \cup B) \cap (A \cup C) \\
A \cup (B \cap \llcorner B) = A & \llcorner(A \cup B) = \llcorner A \cap \llcorner B \\
\llcorner\llcorner A = A & A \cup \llcorner A = H \\
\multicolumn{2}{c}{\llcorner H = \emptyset.}
\end{array}
}
\tag{1}
$$

H-алгебра является частным случаем так называемой *булевой алгебры* (см. «Алгебра»). Вышеперечисленные 9 тождеств вместе с законами замены и дополнения представляют собой одну *систему аксиом* булевой алгебры.

Дуальным выражением H-алгебры считается такое выражение, которое получится из первоначального выражения путем замены знаков $\cup$, $\cap$, H, $\emptyset$ знаками $\cap$, $\cup$, $\emptyset$, H соответственно. (Если в некотором выражении ни один из этих знаков не присутствует, то дуальным выражением будет само выражение.) Например, дуальным выражением выражения $(A \cup B) \cap H$ является $(A \cap B) \cup \emptyset$.

Дуальность. Как в общем виде в булевой алгебре, так и в H-алгебре действителен *закон дуальности:* если образовать дуальные выражения на обеих сторонах некоторого множества (т. е., если поставить дуальное выражение каждого выражения на обеих сторонах), то получится тождество. Например, два закона де Моргана являются дуальными законами.

В тождествах (1) (в «аксиомах») не присутствуют дуальные тождества (за исключением тождества $\llcorner\llcorner A = A$, дуальным тождеством которого является само тождество). На основе тождеств (1) получатся и дуальные тождества.

МНОЖЕСТВА И ФУНКЦИИ

ОБЩЕЕ ПОНЯТИЕ ФУНКЦИИ

Пусть X — не пустое множество, и предположим, что каждому элементу множества X *соответствует именно один* предмет. Обозначим через Y множество предметов, соответствующих элементам множества X. В самом общем смысле слова

функцией называется любое такое правило соответствия, на основе которого каждому элементу множества X соответствует именно по одному предмету.

Для обозначения функций применяются латинские прописные буквы (f, g, h, . . .). Если f представляет собой функцию, на основе которой предметы соответствуют элементам множества X, а множество соответствующих предметов — Y, то говорят, что множество X является областью существования функции f, а Y — множеством значений функции.

Для обозначения вышеприведенного соотношения иногда применяется символ $X \xrightarrow{f} Y$. Элементы X часто называются *аргументами* функции, а элементы Y — *значениями* функции. С помощью f X «преобразуется» в Y, или Y является «изображением» X, созданного с помощью f. Если $x \in X$, то можно обозначить через $f(x)$ предмет, соответствующий значению x с помощью f. Таким образом, с помощью равенства

$$f(x) = y$$

получается, что элемент y множества значений функции f соответствует элементу x области существования функции f.

На основе приведенного выше определения функцией является, например, платежная ведомость, по которой каждому рабочему соответствует определенная заработная плата. Функцией является также и такое правило, по которому каждому положительному числу, больше 1, предписывается самый большой неделимый фактор этого числа, или такое указание, по которому каждому действительному числу предписывается двойное значение этого числа и т. д.

Особое значение имеют те функции, у которых и область существования, и множество значений функции представляют собой некоторое числовое множество. Как известно, с помощью таких функций выражаются *изменения* природы в математической форме.

Замечание. На основе приведенного выше понятия функции не предписывается (но и не запрещается), то, что разным элементам области существования некоторой функции обязательно должны соответствовать разные значения функции. Множества областей существования и значений функции не обязательно должны представлять дизъюнктивные, либо различные множества.

Естественно, такое понятие функции является довольно общим; оно не ограничено ни тем, что элементы области существования и значений функции должны быть числами, ни тем, чтобы было возможно выразить единственной формулой правило предписания.

Функция всегда может быть представлена множеством таких упорядоченных пар, первые члены которых являются элементами области существования функции, а вторые члены — значениями функции, принадлежащими соответствующим первым членам. Таким образом, множество A упорядоченных пар представляет функцию $x \xrightarrow{f} Y$, если $(x, y) \in A$ тогда и только тогда, если $x \in X$, и $y = f(x)$. Наоборот, можно считать множество A упорядоченных пар заданием функции f, если нет таких двух различных пар среди его элементов, первые члены которых тождественны. Следовательно, во множестве нет такой пары (x, y_1) и (x, y_2), относительно которых $y_1 \neq y_2$, т. е., одному значению x соответствует только одно значение y. В таком случае множество первых членов пар в множестве A является областью существования f, а множество вторых членов пар — множеством значений функции f, и

$$f(x) = y \quad \text{тогда и толко тогда, если} \quad (x, y) \in A.$$

Замечание. Когда функция задается таблицей значений, то двумя строками таблицы значений выражается именно множество упорядоченных пар : элементы, записанные друг под другом (x и y соответственно) образуют упорядоченные пары.

СТЕПЕНЬ, ОБРАЗОВАННАЯ ИЗ МНОЖЕСТВ

Пусть A и B — непустые множества. Под понятием степени множества A с экспонентом B подразумевается множество всех функций, областью существования которых является B, а значениями — элементы множества A. Такое множество обозначается символом A^B.

Если, **например,** множество действительных чисел обозначается буквой V, то V^V представляет собой множество всех действительных функций. Пусть

$$A = \{0, 1, 2\}\ ; \quad B = \{a, b\}\ .$$

В множестве B могут быть определены 9 таких функций, значения которых выбираются из элементов множества A. Эти функции (с обозначениями $f_1, f_2, \ldots f_8, f_9$ соответственно) задаются следующей таблицей значений :

		f_1	f_2	f_3	f_4	f_5	f_6	f_7	f_8	f_9
B	a	0	0	0	1	1	1	2	2	2
	b	0	1	2	0	1	2	0	1	2

$$f_1(a) = 0, \quad f_1(b) = 0\ ; \quad f_2(a) = 0, \quad f_2(b) = 1\ ; \ldots$$

И, таким образом,

$$A^B = \{f_1, f_2, \ldots, f_8, f_9\}\ .$$

Этот пример показывает, что

количество элементов $A^B =$ (количество элементов $A)^{(\text{количество элементов } B)}$.

Можно доказать, что если A и B являются ограниченными и непустыми множествами, то вышеприведенное утверждение всегда правильно и этим объясняется обозначение A^B и название данной операции.

Если в частном случае $A = \{0, 1\}$, то элементами A^B (т. е. элементами $\{0, 1\}^B$) являются все функции, определенные в множестве B, значения которых могут быть исключительно числа 0 и 1.

Пусть f представляет собой такую функцию, а C — подмножество множества B, относительно элементов которого функция f принимает значение 1. Функция f называется характеристической функцией (относительно B) множества C. Такая функция характеризуется тем, что

$$f(x) = \begin{cases} 1, & \text{если} \quad x \in C\,, \\ 0 & \text{иначе.} \end{cases}$$

Очевидно, что каждая часть множества B однозначно определяется такой характеристической функцией; $\{0, 1\}^B$ содержит в себе именно характеристические функции (относительно множества B) частей B. Таким образом, множеством $\{0, 1\}^B$ также однозначно (посредственно) задаются части B, как и множеством 2^B.

ЭКВИВАЛЕНТНОСТЬ МНОЖЕСТВ

Функция f называется *обратимой* (инвертируемой), если в ней разным элементам области существования предписаны разные значения функции. *Обратной функцией* некоторой обратимой функции $x \xrightarrow{f} y$ называется функция $Y \xrightarrow{g} X$, для которой выполняется следующее условие:

$g(y) =$ (такому значению x, для которого $f(x) = y$).

(Существует единственное такое значение x, так как функцией f задаются разные значения y для разных значений x.) Если функция f является обратимой, то тогда обратная функция задается однозначно; обратная функция некоторой функции f обычно обозначается символом $\bar{f}$.

Примеры: Если функцией f предписывается каждому женатому мужчине его жена, тогда функцией $\bar{f}$ предписывается каждой замужней женщине ее муж (если предполагать моногамию и моноандрию!). Если функцией g предписывается каждому действительному числу его половина, то функцией $\bar{g}$ предписывается каждому действительному числу его двойное значение.

Областью существования и множеством значений некоторой обратимой функции определяется так называемое *взаимно однозначное предписание* (*предписание* типа 1—1): каждому элементу области существования обратимой функции предписывается именно единственный элемент множества значений функции, и любой элемент множества значений предписывается именно единственному элементу области существования функции (так как в таких функциях разные значения предписываются разным значениям). В сущности, обратной функцией осуществляется такое же предписание, только наоборот.

Два множества называются *эквивалентными*, если между их элементам устанавливается предписание типа 1—1. Иными словами: два множества являются эквивалентными, если существует такая обратимая функция, областью существования которой является одно, а множеством значений функции является другое множество. Эквивалентность множеств A и B обозначается символом «$A \sim B$».

Замечание. Если A является множеством, то функцией, с помощью которой любому элементу A предписывается сам элемент, осуществляется предписание типа 1—1 между элементами A и элементами A. Такое рассуждение является в том случае правильным, (хотя и звучит несколько странно), если A представляет собой пустое множество. Таким образом, каждое множество, включая и пустое множество, оказывается эквивалентным самому себе.

Из определения эквивалентности непосредственно вытекают следующие свойства:

(а) Для любого множества: A: $A \sim A$ *(рефлексивность).*
(б) Если $A \sim B$, то $B \sim A$ *(симметрия).*

Следующие свойства приводятся более сложным путем:

(в) Если $A \sim B$, и $B \sim C$, то $A \sim C$.
(г) Если $A \sim A'$, $B \sim B'$, далее A и B являются дизъюнктивными множествами, A' и B' — также дизъюнктивными, то $A \cup B \sim A' \cup B'$.
(д) Если $A \sim A'$ и $B \sim B'$, то $A \times B \sim A' \times B'$ и $A^B \sim A'^{B'}$.
(е) Если $A \subseteq B$ и $B \sim B'$, то существует такое множество A', что $A' \subseteq B'$, и $A \sim A'$.
(ж) Для любого множества A: $2^A \sim \{0,1\}^A$.

Тот факт, что ограниченное множество имеет n элементов (где n — положительное целое число), в действительности означает, что данное множество является эквивалентным множеству $\{1,2, \ldots, n-1, n\}$. В действительности элементы ограниченного множества подсчитываются, когда осуществляется предписание типа 1—1 между элементами множества и положительными целыми числами, применяемыми при расчете. Таким образом, если учесть свойства эквивалентности, то можно сделать вывод: *два ограниченных множества могут быть эквивалентными тогда и только тогда, если числа их элементов равны.*

Следствием вышеуказанного является следующее:

Ограниченное множество никогда не может быть эквивалентным его некоторой действительной части.

ИСЧИСЛИМЫЕ МНОЖЕСТВА

ИСЧИСЛИМОЕ МНОЖЕСТВО И БЕСКОНЕЧНАЯ ПРОГРЕССИЯ

Основной чертой теории множества является то, что с ее помощью различаются бесконечности различных масштабов и возможны расчеты с бесконечностями. Самым простым примером бесконечного множества является множество естественных чисел. У этого множества существуют и *действительные* части, эквивалентные ему. Ниже задаем несколько обратимых функций (вместе с их обратными функциями), определенных в множестве естественных чисел, множество значений которых является действительным подмножеством естественных чисел. (В формулах n представляет собой любое естественное число — 0, 1, 2, 3, . . . Следовательно, 0 считается здесь также естественным числом.)

Функция	Обратная функция
$f(n) = n + 1$	$\bar{f}(n + 1) = n$
$g(n) = 2 \cdot n$	$\bar{g}(2 \cdot n) = n$
$h(n) = 2 \cdot n + 1$	$\bar{h}(2 \cdot n + 1) = n$
$k(n) = 10^n$	$\bar{k}(10^n) = n$

В первой строке приведенной ниже таблицы представлены естественные числа, а в последующих — значения функций f, g, h, k соответственно. Из таблицы видно, что между множеством естественных чисел и некоторыми его действительными частями (положительные целые, четные и нечетные естественные числа, десятичные степени) могут быть осуществлены предписания типа 1—1.

Естественные числа:	0	1	2	3	4	5	. . .	n	. . .
Положительные целые, $f(n)$;	1	2	3	4	5	6	. . .	$n+1$	. . .
Четные числа; $g(n)$:	0	2	4	6	8	10	. . .	$2 \cdot n$	. . .
Нечетные числа; $h(n)$:	1	3	5	7	9	11	. . .	$2 \cdot n+1$	. . .
Десятичные степени; $k(n)$:	10^0	10^1	10^2	10^3	10^4	10^5	. . .	10^n	. . .

На основе примеров читатель самостоятельно может назвать несколько действительных частей множества есетественных чисел, эквивалентных ему.

Множества, эквивалентные множеству естественных чисел, называются исчислимыми множествами.

(Некоторые авторы называют исчислимыми и ограниченные множества.)

Бесконечная прогрессия. Функции (бесконечные), определенные в множестве естественных чисел, называются *прогрессиями.* Если обозначить через a_n значение функции, предписанное есетственному числу n, то предметы

$$a_0,\ a_1,\ a_2,\ a_3,\ a_4,\ \ldots,\ a_n,\ \ldots$$

называются *членами* прогрессии. На основе определения прогрессии бесконечное множество является исчислимым в том и только в том случае, если его элементы «могут быть упорядочены как бесконечная прогрессия». Но, в то же время, множество членов некоторой прогрессии не всегда является бесконечным множеством, так как члены прогрессии как члены функции не различаются в любом случае (пример: каждый член прогрессии является одним и тем же предметом); но если множество является бесконечным, то в любом случае исчислимо.

Если пропустить или прибавить один член в начале прогрессии, сама прогрессия остается бесконечной. Этот процесс можно повторять. Из этого следует: Путем соединения ограниченного и исчислимого множеств получается всегда исчислимое множество. Множество, полученное из исчислимого множества путем пропуска ограниченного количества элементов, является также исчислимым множеством.

Можно пропускать и бесконечное количество членов из некоторой прогрессии (например, каждый второй член), и если при этом остается бесконечное количество членов, то и в этом случае эти члены представляют собой бесконечную прогрессию. Поэтому:

Любая бесконечная часть исчислимого множества является исчислимой. Иными словами: каждая часть исчислимого множества является или ограниченной, или исчислимой.

Две бесконечных прогрессии всегда могут быть соединены в одну бесконечную прогрессию. Например, прогрессии

$$a_0, a_1, a_2, \ldots, a_n, \ldots \qquad b_0, b_1, b_2, \ldots, b_n, \ldots$$

могут быть соединены в прогрессии

$$a_0, b_0, a_1, b_1, a_2, b_2, \ldots, a_n, b_n, \ldots$$

Путем повторения данного процесса получится следующее:

При соединении ограниченного количества исчислимых множеств получится также исчислимое множество.

Процесс соединения может быть распространен на случай, когда количество членов бесконечно. Под определением либо конечного, либо бесконечного количества множеств подразумевается множество таких предметов, которые являются элементом хотя бы одного из соединимых множеств.

Легко убедиться в том, что в результате соединения исчислимого бесконечного количества ограниченных множеств получится либо конечное, либо исчислимое множество (в крайнем случае исчислимое). Множество будет наверняка исчислимо, если соединяемые множества являются дизъюнктивными по парам.

ХАРАКТЕРНЫЕ СВОЙСТВА БЕСКОНЕЧНЫХ МНОЖЕСТВ

На основе сказанного выше легко убедиться в том, что исчислимые множества являются «самыми простыми» бесконечными множествами. Подобное рассуждение станет более понятным, если ввести следующую теорему:

Каждое бесконечное множество имеет исчислимую часть. Это справедливо потому, что из некоторого бесконечного множества может быть выделена следующая прогрессия:

$$a_0, a_1, a_2, \ldots, a_n, \ldots,$$

которая состоит исключительно из разных элементов. Во всяком случае, из множества может быть выбран элемент a_0 (так как множество не является пустым), затем выбор может продолжаться после каждого шага. Множество не может стать пустым, так как в противном случае, если после миллионного шага оно стало бы пустым, то оно имело бы только конечное количество — 1 миллион членов, вопреки тому, что данное множество является бесконечным.

Мы убедились, что любое исчислимое множество имеет эквивалентную ему действительную часть. Из этого и из предыдущей теоремы следует: *каждое бесконечное множество имеет эквивалентную ему действительную часть.*

Если множество A является бесконечным, то на основе предыдущей теоремы такое множество может быть воспроизведено в виде $M\cup R$, где множество M исчислимо и $R = A\setminus M$; таким образом, M и R являются дизъюнктивными и $A = M\cup R$. Но множество M имеет такую действительную часть M', что $M\sim M'$; очевидно, что M' и R являются дизьюнктивными. Таким образом,

$$\left.\begin{array}{r} M\sim M' \\ R\sim R \end{array}\right\} \rightarrow \underbrace{M\cup R}_{A} \sim \underbrace{M'\cup R}_{A'}$$

и A' является действительной частью множества A.

Поскольку конечное множество никогда не имеет эквивалентную ему действительную часть, то на основе последней теоремы формулируется следующее определение конечных и бесконечных множеств (в этом определении мы уже не ссылаемся на естественные числа и на расчеты):

множество считается конечным, если не имеет эквивалентную ему действительную часть. Множество считается бесконечным, если оно не является конечным (т. е. если имеет эквивалентную ему действительную часть).

ДАЛЬНЕЙШИЕ ИСЧИСЛИМЫЕ МНОЖЕСТВА

Множество целых чисел является соединением двух исчислимых множеств (множества естественных чисел и множества отрицательных целых чисел), поэтому оно также является исчислимым. Множество положительных дробей может быть воспроизведено путем соединения исчислимого количества исчислимых множеств на основе приведенной ниже таблицы.

$\frac{1}{1}$	$\frac{2}{1}$	$\frac{3}{1}$	$\frac{4}{1}$	$\frac{5}{1}$	...
$\frac{1}{2}$	$\frac{2}{2}$	$\frac{3}{2}$	$\frac{4}{2}$	$\frac{5}{2}$	...
$\frac{1}{3}$	$\frac{2}{3}$	$\frac{3}{3}$	$\frac{4}{3}$	$\frac{5}{3}$	...
$\frac{1}{4}$	$\frac{2}{4}$	$\frac{3}{4}$	$\frac{4}{4}$	$\frac{5}{4}$	...
$\frac{1}{5}$	$\frac{2}{5}$	$\frac{3}{5}$	$\frac{4}{5}$	$\frac{5}{5}$	...
...	...	...	...	...	...

В то же время настоящая таблица может быть «выпрямлена» по изображенному диагональному пути в виде единственной бесконечной прогрессии и получится на первый взгляд удивительный факт, что *множество положительных дробей* является также «только» *исчислимым.* (В таблице каждая дробь представлена в бесконечно многих видах. Если из дробей, одинаковых по значению, оставить только одну (например, ту, которую мы нашли впервые), то получится также исчислимое множество, так как каждая бесконечная часть исчислимого множества является исчислимой. Бесспорно, что оставшееся множество является бесконечным, так как его элементами являются все положительные целые числа (как дроби с единичным знаменателем). Из этого следует:

Путем соединения исчислимо-бесконечного количества исчислимых множеств получится также исчислимое множество.

Множество отрицательных дробей, очевидно, является эквивалентным множеству положительных дробей, поэтому также исчислимо. Если соединить эти два множества, а затем прибавить к результату элемент 0, то получится также исчислимое множество. Таким образом:

множество рациональных чисел также исчислимо.

Так как положительные дроби могут считаться *упорядоченными парами*, образованными из положительных целых чисел, то из вышеуказанного диагонального метода следует, что

множество M_2 упорядоченных пар, образуемых из элементов некоторого исчислимого множества M, также исчислимо.

Упорядоченные тройственные числа типа (a, b, c) можно рассматривать как упорядоченную пару типа $((a, b), c)$, первым членом которой является упорядоченная пара (a, b), а вторым членом — c. Таким образом, множество тройственных чисел, образуемых из элементов некоторого исчислимого множества M, является определенной (бесконечной) частью множества упорядоченных пар, образуемых из элементов исчислимого множества $M_2 \cup M$ (так как в эту часть входят только те пары, первый член которых принадлежит множеству M_2, а второй член — множеству M; здесь множество M_2 является опять множеством упорядоченных пар, образуемых из элементов множества M); поэтому оно также исчислимо.

Примеры к последним двум выводам: сетевые точки плоскости, образуемой из единичных квадратов, и сетевые точки пространства, образуемого из единичных кирпичей, составляют исчислимое множество. Путем повторения данной операции получится, что *множество упорядоченных* 4-, 5-, ..., *n-ых чисел, образуемых из элементов некоторого исчислимого множества, является также исчислимым.* Рациональный многочлен

$$a_0 + a_1x + a_2x^2 + \ldots + a_nx^n \qquad (n \geqq 1)$$

с целыми коэффициентами определяется однозначно *упорядоченными* $(n + 1)$-*числами* формы $(a_0, a_1, a_2, \ldots, a_n)$, образованными из коэффициентов. Таким образом, множество многочленов с рациональными целыми коэффициентами максимум n-го порядка является также исчислимым.

Множество всех рациональных многочленов с целыми коэффициентами может быть воспроизведено путем соединения множеств рациональных многочленов с целыми коэффициентами максимум первого, второго, ..., n-го порядка, т. е. путем соединения неисчислимо-бесконечного количества исчислимых множеств. Таким образом, данное множество также является исчислимым.

Алгебраические числа (см. гл. «Теория чисел») являются корнями рациональных многочленов с целыми коэффициентами. У такого многочлена может быть в любом случае только конечное количество различных корней (см. гл. «Алгебра».) Поскольку множество таких многочленов является исчислимым, то множеством их корней является объединение исчислимо-бесконечного количества конечных множеств. Следовательно, множество (либо действительных, либо сложных) алгебраических чисел является по крайней мере исчислимым. Так как данное множество содержит в себе, например, целые числа, то оно наверняка является бесконечным и, следовательно, однозначно исчислимым.

НЕИСЧИСЛИМЫЕ МНОЖЕСТВА

МНОЖЕСТВА НЕПРЕРЫВНОЙ ЧИСЛЕННОСТИ

На основе сказанного выше у читателя может возникнуть предположение, что каждое бесконечное множество является исчислимым. Но это не так. На примере будет показано, что множество действительных чисел между границами 0 и 1 (включая 0, а 1 не включая), т. е. *интервал* $[0\,;\ 1]$, закрытый слева и открытый справа, является неисчислимым (см. гл. «Анализ»).

Действительное число, которое является элементом данного множества, может быть записано в виде бесконечной десятичной дроби, начинающейся с 0:

$$0,a_1a_2a_3\ldots a_n\ldots .$$

Здесь $a_1, a_2, a_3, \ldots, a_n, \ldots$ обозначают одно из чисел 0, 1, 2, ..., 8, 9 с тем ограничением, что в данной прогрессии не могут стоять — начиная с некоторого знака — одни знаки 9. Такое ограничение необходимо, так как бесконечные десятичные дроби такой формы могут быть выражены в виде конечных десятичных дробей:

$$0,372\,999\ldots = 0,373$$

(последний, отличающийся от 9, знак увеличивается на 1 и бесконечная часть, состоящая из одних знаков 9, отставляется). В то же время конечные десятичные дроби, дополненные знаками «0» (например, 0,373 000 ...), уже присутствовали среди бесконечных десятичных дробей. Таким образом, цель ограничения: бесконечные десятичные дроби не должны быть в двух видах. При соблюдении данного ограничения каждое действительное число, принадлежащее вышеуказанному множеству, выражается именно одной бесконечной десятичной дробью, начинающейся знаком «0».

Таким образом, в дальнейшем следует показать, что множество таких бесконечных десятичных дробей является неисчислимым, то есть оно не может быть упорядочено в прогрессию.

Такая прогрессия имеет следующий вид:

$$\begin{aligned}
\alpha_1 &= 0,\, a_{11}a_{12}a_{13} \ldots a_{1n} \ldots \\
\alpha_2 &= 0,\, a_{21}a_{22}a_{23} \ldots a_{2n} \ldots \\
\alpha_3 &= 0,\, a_{31}a_{32}a_{33} \ldots a_{3n} \ldots \\
&\ldots\ldots\ldots\ldots\ldots\ldots \\
\alpha_n &= 0,\, a_{n1}a_{n2}a_{n3} \ldots a_{nn} \ldots \\
&\ldots\ldots\ldots\ldots\ldots\ldots
\end{aligned}$$

Но к каждой такой прогрессии может быть составлена такая бесконечная десятичная дробь, начинающаяся с 0 целых, которая соответствует данному ограничению, но не является членом вышеуказанной прогрессии. Для этой цели выбираем численные знаки $b_1, b_2, b_3, \ldots, b_n, \ldots$ следующим образом:

b_1 отличается от 9 и от значения a_{11},

b_2 отличается от 9 и от значения a_{22},

b_3 отличается от 9 и от значения a_{33},

. . . .

b_n отличается от 9 и от значения a_{nn}

. . . .

Каждый из этих численных знаков выбирается свободно из восьми (а может быть, из девяти) численных знаков. Образованная из них бесконечная десятичная дробь

$$\beta = 0,\, b_1b_2b_3 \ldots b_n \ldots$$

удовлетворяет вышеуказанному ограничению (так как знак 9 в ней вовсе не присутствует) и отличается от каждого члена приведенной выше прогрессии. Эта дробь наверняка отличается от α_1 в первом десятичном знаке (поскольку $b_1 \neq a_{11}$), от α_2 во втором десятичном знаке (так как $b_2 \neq a_{22}$), от α_3 в третьем десятичном знаке ($b_3 \neq a_{33}$), . . ., от α_n в n-ом десятичном знаке ($b_n \neq a_{nn}$) и т. д. Таким образом, эта дробь наверняка не является членом вышеуказанной прогрессии. (Естественно, это не является единственной такой бесконечной десятичной дробью.)

Следовательно, невозможно составить такую прогрессию, которая содержала бы в себе *все* бесконечные десятичные дроби, начинающиеся с 0 целых (удовлетворяющие вышеуказанному ограничению), т. е. содержала бы все действительные числа между границами 0 и 1. Из этого вытекает, что данное множество не может быть исчислимым.

Нетрудно убедиться в том (на основе соответствующих теорем, относящихся к исчислимым множествам, и на основе того, что каждое бесконечное множество имеет исчислимую часть), что путем соединения любого бесконечного множества V и исчислимого множества всегда получается множество, эквивалентное множеству V.

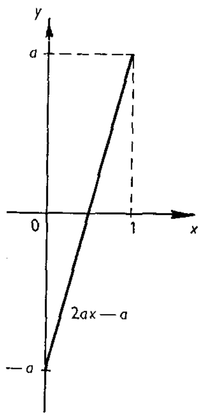
Из этого и из предыдущего доказательства следует, что

множество действительных чисел между границами 0 *и* 1 *является неисчислимым*

(независимо от того, включаем ли мы одно из чисел 0 или 1, либо оба числа, либо не включаем ни одного). Эквивалентные этому множества называются множествами *непрерывной численности*. (Слово «численность» поясняется ниже.)

Если a является положительным действительным числом, то обратимая функция

$$f(x) = 2ax - a \qquad (0 \leqq x \leqq 1)$$

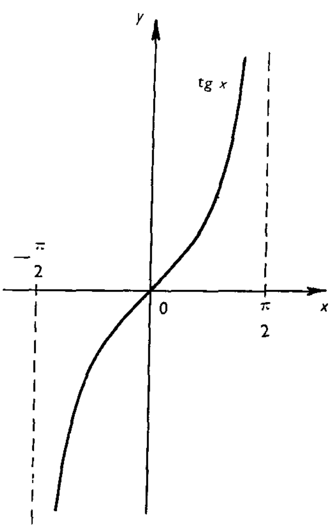


(обратная функция:

$$\bar{f}(y) = \frac{y+a}{2a}; \quad -a \leqq y \leqq a)$$

осуществляет предписание типа 1—1 между интервалами $[0;\ 1]$ и $[-a;\ a]$ (см. рис.). Таким образом, обратная функция является также и множеством непрерывной численности. Если к сказанному выше добавить, что два интервала одинаковой длительности являются всегда эквивалентными, то получится следующее:

любой конечный интервал (независимо от того, включены ли конечные точки в интервал) *действительных чисел является интервалом непрерывной численности.*



Функция $f(x) = \operatorname{tg} x$ на интервале $\left(-\frac{\pi}{2};\ \frac{\pi}{2}\right)$ также является обратимой (см. гл. «Анализ»), таким образом этой функцией осуществляется предписание типа 1—1

а) между интервалами $\left(0;\ \frac{\pi}{2}\right)$ и $[0;\ +\infty)$,

б) между интервалами $\left(\frac{-\pi}{2};\ \frac{\pi}{2}\right)$ и $(-\infty;\ +\infty)$. Из этого и нижеследующего следует: *любой бесконечный интервал действительных чисел и, соответственно, множество всех действительных чисел является интервалом непрерывной численности.*

Если исчислимое множество рациональных чисел соединяется с бесконечным множеством иррациональных чисел, то получится множество, эквивалентное множеству иррациональных чисел (см. выше). С другой стороны, результатом соединения является множество действительных чисел. Таким образом, получится, что

множество иррациональных чисел является множеством непрерывной численности.

Подобным образом получится, что

множество трансцендентных (действительных) чисел (см. гл. «Теория чисел») *также является множеством непрерывной численности.*

Между точками прямой и действительными числами может быть установлено соответствие типа 1—1; на этом основано изображение действительных чисел по численной прямой. Конечным интервалам действительных чисел соответствуют отрезки, а бесконечным с одной стороны интервалам — полупрямые. Следовательно, множество точек любого отрезка, любой полупрямой или прямой является множеством непрерывной численности.

Может быть доказано, что такие множества, как множество бесконечных прогрессий упорядоченных пар, трехзначных, ..., n-значных, ... чисел, образуемых из элементов множества непрерывной численности, множество *непрерывных* действительных функций (см. гл. «Анализ»). множество комплексных чисел, степенное множество некоторого исчисчислимого множества и т. д. являются также множествами непрерывной численности.

Более важное значение имеют множества непрерывной численности в анализе действительных и комплексных функций, а также и в других областях математики.

ЧИСЛЕННОСТИ

Как оказалось, не все бесконечные множества являются исчислимыми, поэтому имеет смысл обобщить понятие чисел таким образом, чтобы данное понятие можно было применять при «измерении» бесконечных множеств. Для этой цели существует понятие *численности.*

Каждому множеству предписывается понятие, названное численностью данного множества (или количественным числом множества) с тем условием, что численности эквивалентных множеств одинаковы, а численности неэквивалентных множеств различаются.

Численность множества A обозначается символом $\overline{\overline{A}}$. Если $\overline{\overline{A}} = \overline{\overline{B}}$, то это имеет такое же значение, как выражение $A \sim B$.

Так как два конечных множества являются эквивалентными тогда и только тогда, когда они имеют одинаковое количество элементов, то численность некоторого конечного множества обозначается числом элементов множества. Таким образом, *естественные числа являются численностями конечных множеств;* они называются конечными численностями, а численности бесконечных множеств называются бесконечными численностями. $\overline{\overline{\emptyset}} = 0$ и если конечное множество A имеет n элементов, то $\overline{\overline{A}} = n$. Численности исчислимых множеств обозначаются через **a**, а численности множеств непрерывной численности — через **c**. (Для самостоятельного обозначения численностей часто применяются готические прописные буквы. Для простоты мы обозначаем численности жирным шрифтом.)

(Численная прямая заполняется действительными числами без пропусков, непрерывно. На основе введения понятия о численности становится понятным присутствие слова «численности» в понятии «множество непрерывной численности». Исчислимые множества можно также называть и «множествами исчислимой численности».)

Множества численности **a** «больше», чем любое конечное множество численности n, в том смысле, что они имеют часть численности n, но в то же время они сами не являются множествами численности n. Таким же образом множество численности **c** «больше», чем множества численности **a**, в том смысле, что они имеют часть численности **a**, а сами не являются множествами численности **a**.

Из сказанного выше следует, что любую конечную численность n целесообразно рассматривать как *меньшую*, чем **a**, а численность **a** — *меньшую*, чем **c**. (С этой целью введены понятия «меньше» и «больше» для круга численностей.)

Говорят, что численность $\mathfrak{m}$ *меньше*, чем численность $\mathfrak{n}$ (или также: $\mathfrak{n}$ *больше*, чем $\mathfrak{m}$), если множество M численности $\mathfrak{m}$ является эквивалентным некоторой части множества N численности $\mathfrak{n}$, но не является эквивалентным самому множеству N; это обозначается символом «$\mathfrak{m} < \mathfrak{n}$» (или «$\mathfrak{n} > \mathfrak{m}$»).

Замечания: 1. Безразлично, какое из множеств численностей $\mathfrak{m}$ и $\mathfrak{n}$ выбирается как M и N в соответствии с определением эквивалентности, представленным на стр. 553, в частности, в соответствии с условиями б), в) и е); если после выбора множеств M и N получится $\mathfrak{m} < \mathfrak{n}$, то при другом их выборе получится такое же неравенство.

2. Для конечных численностей $\mathfrak{m}$ и $\mathfrak{n}$, то есть для естественных чисел, условие $\mathfrak{m} < \mathfrak{n}$ справедливо в том и только в том случае, если $\mathfrak{m}$ является меньшим естественным числом, чем $\mathfrak{n}$ (понятия «меньше» в том смысле, в котором оно применяется в области естественных чисел). Действительно, если M и N являются конечными множествами и множество M эквивалентно некоторой части множества N, но не является эквивалентным множеству N, то M может быть эквивалентно только некоторой *действительной* части множества N и, таким образом, множество M имеет меньше элементов, чем множество N.

3. Если множества M и N являются бесконечными и множество M эквивалентно некоторой *действительной* части множества N, то не исключается правильность соотношения $M \sim N$ (такие примеры были представлены выше). Поэтому не было бы достаточным условием в определении соотношения «меньше», что множество M эквивалентно некоторой *действительной* части множества N.

Соотношение «меньше» для естественных чисел характеризуется следующими свойствами. Если a, b, c являются естественными числами, то:

1. $a < a$ исключено	*(иррефлексивность)*;
2. при $a < b$ исключено: $b < a$	*(асимметрия)*;
3. если $a < b$ и $b < c$, то из этого следует, что $a < c$	*(транзитивность)*;
4. при $a \neq b$ выполняется одно из следующих двух условий: $a < b$ и $b < a$	*(трихотомия)*.

Возникает вопрос: действительны ли эти свойства для *бесконечных* численностей? Если $\mathfrak{m}$ является численностью, то по определению соотношение $\mathfrak{m} < \mathfrak{m}$ означало бы, что множество M (численности $\mathfrak{m}$) не было бы эквивалентно самому себе, что невозможно. Следовательно, *иррефлексивность* соотношения «$<$» в общем случае действительна для численностей.

При одновременном выполнении соотношений $\mathfrak{m} < \mathfrak{n}$ и $\mathfrak{n} < \mathfrak{m}$ существовали бы такие множества M и N, которые взаимно эквивалентны части друг друга, но не эквивалентны между собой. Но на основе так называемой *теоремы эквивалентностей* из выполнения условий $M \sim N'$, $N' \subseteq N$ и $N \sim M'$, $M' \subseteq M$, следует выполнение соотношения $M \sim N$. (Доказательство данной теоремы здесь не представляется.) Таким образом, асимметрия соотношения «$<$» действительна в общем случае для численностей.

На основе этих двух заключений непосредственно следует *транзитивность* соотношения «$<$» для численностей.

Наконец, если бы соотношение «$<$» для численностей не было бы трихотомным, то это означало бы, что существуют «несравнимые» между собой пары множеств, т. е. такие пары множеств M, N, что множество M не эквивалентно части множества N, а множество N не эквивалентно части множества M. Только в области теории упорядоченных множеств удалось доказать (см. ниже), что такие пары не могут существовать, т. е., что любые два множества могут «сравниваться» между собой.

Операции численностями. В области численностей могут быть определены также и операции сложения, умножения и возведения в степень. Эти определения (подобно определению соотношения «$<$») представлены в таком виде, что если применять их для конечных численностей, то есть, для естественных чисел, то получатся соответствующие операции, известные для естественных чисел.

Сумма двух численностей $\mathfrak{m}$ и $\mathfrak{n}$ обозначается символом $\mathfrak{m} + \mathfrak{n}$ и определяется следующим образом. Пусть M и N будут такими дизъюнктивными множествами, для которых $\overline{\overline{M}} = \mathfrak{m}$, $\overline{\overline{N}} = \mathfrak{n}$. В таком случае

$$\mathfrak{m} + \mathfrak{n} = \overline{\overline{M \cup N}}.$$

Произведение численностей $\mathbf{m}$ и $\mathbf{n}$ обозначается символом $\mathbf{m} \cdot \mathbf{n}$ и представляется в следующем виде. Пусть M и N — такие множества, для которых $\overline{\overline{M}} = \mathbf{m}, \overline{\overline{N}} = \mathbf{n}$. Тогда

$$\mathbf{m} \cdot \mathbf{n} = \overline{\overline{M}} \times \overline{\overline{N}}.$$

Возведение в степень $\mathbf{m}^{\mathbf{n}}$, образуемое из численностей, определяется следующим образом:

$$\mathbf{m}^0 = 1 \text{ (для любого } \mathbf{m}),$$
$$0^{\mathbf{n}} = 0 \text{ если } \mathbf{n} \neq 0,$$

а если и $\mathbf{m}$ и $\mathbf{n}$ отличаются от 0, то пусть будут M и N, для которых $\overline{\overline{M}} = \mathbf{m}$ и $\overline{\overline{N}} = \mathbf{n}$. Определение в этом случае имеет следующий вид:

$$\mathbf{m}^{\mathbf{n}} = \overline{\overline{M^N}}.$$

Результат операций не зависит от выбора множеств M и N, присутствующих в определении. Это обеспечивается свойствами г) и д) эквивалентностей (см. стр. 553). По замечаниям, сделанным после определений соотношений союза, непосредственного произведения и возведения в степень множеств (см. стр. 541, 546, 551), эти операции являются обобщениями операций соответствующих названий, определенных для естественных чисел. Можно доказать, что основные тождества операций также остаются действительными для обобщенных операций: сложение и произведение являются коммутативным и ассоциативным, произведение является дистрибутивным для сложения, а для возведения в степень действительны следующие соотношения:

$$(\mathbf{m} \cdot \mathbf{n})^{\mathbf{r}} = \mathbf{m}^{\mathbf{r}} \cdot \mathbf{n}^{\mathbf{r}}, \quad \mathbf{m}^{\mathbf{n}} \cdot \mathbf{m}^{\mathbf{r}} = \mathbf{m}^{\mathbf{n}+\mathbf{r}}; \quad (\mathbf{m}^{\mathbf{n}})^{\mathbf{r}} = \mathbf{m}^{\mathbf{n} \cdot \mathbf{r}}.$$

На основе применения установлений, сделанных в первом абзаце главы «Алгебра степенных множеств» и свойства ж) эквивалентности (см. стр. 553), очевидно, что для любого множества M действительно следующее соотношение:

$$\overline{\overline{2^M}} = 2^{\overline{\overline{M}}}.$$

Далее можно доказать, что если по крайней мере одна из двух численностей является бесконечной, то их сумма и (если меньшая численность не равняется 0) произведение равняется большей численности.*

В отношении возведения в степень существуют следующие соотношения для численностей $\mathbf{a}$ и $\mathbf{c}$:
Если $\mathbf{m}$ является конечной численностью и она больше 1, то $\mathbf{m}^{\mathbf{a}} = \mathbf{c}$, $\mathbf{a}^{\mathbf{m}} = \mathbf{a}$, $\mathbf{c}^{\mathbf{m}} = \mathbf{c}$; $\quad \mathbf{a}^{\mathbf{a}} = \mathbf{c}$, $\mathbf{c}^{\mathbf{a}} = \mathbf{c}$.

* В то же время для доказательства этих соотношений необходимо применять теорему о сравнимости численностей, то есть для того соотношения, по которому соотношение «$<$» является трихотомным для численностей.

В частности $2^{\mathfrak{a}} = \mathfrak{c}$. Это равенство показывает, что степенное множество исчислимого множества представляет собой множество непрерывной численности.

Вопрос, что существует ли численность между численностями $\mathfrak{a}$ и $\mathfrak{c}$, т. е. такая численность $\mathfrak{b}$, для которой $\mathfrak{a} < \mathfrak{b} < \mathfrak{c}$, остается открытым, это так называемая *проблема непрерывности*. В дальнейшем мы еще вернемся к этой теме.

ВЫСШИЕ ЧИСЛЕННОСТИ

Существует ли численность больше численности $\mathfrak{c}$? Ответ на данный вопрос положительный: *степенное множество всегда большей численности, чем само множество*. На самом деле, любое множество H эквивалентно некоторой действительной части своего степенного множества, например, множеству его частей с единственным элементом (если предписать подмножество $\{h\}$ каждому элементу h множества H, осуществляется предписание типа 1—1 между элементами множества H и частями с одним элементом H). Но, с другой стороны, H и 2^H не могут быть эквивалентными. Это доказывается следующим образом. Если бы существовало некоторое предписание типа 1—1 между элементами и частями множества H, то надо взять такое множество G элементов H, которые не являются элементами предписанного к ним подмножества. Пусть g будет тем элементом, которому предписанием типа 1—1 предписывается именно множество G. Выполнение соотношения $g \in G$ невозможно, так как в множестве G существуют исключительно такие элементы, которые не являются элементами предписанного подмножества (но, с другой стороны, элементу g предписывается множество G). Также невозможно выполнение соотношения $g \notin G$, потому что это означало бы, что g не является элементом подмножества, предписанного ему, в то время как множество G содержит каждый такой элемент. Соотношения $g \in G$ и $g \notin G$ не могут быть выполнены, так как одно из двух соотношений должно быть правильным; третьего случая нет. Предположение $H \sim 2^H$ неверно.

Таким образом, для всех множеств H выполняется соотношение $\overline{\overline{H}} < \overline{\overline{2^H}}$.

Поэтому $\mathfrak{c} < 2^{\mathfrak{c}}$. Множество всех действительных (определенных на действительных числах, имеющих действительные значения) функций является конкретным примером множества численности $2^{\mathfrak{c}}$.

С помощью введения следующих сокращенных форм

$$2^{\mathfrak{c}} = \mathfrak{c}_1,\ 2^{\mathfrak{c}_1} = \mathfrak{c}_2,\ 2^{\mathfrak{c}_2} = \mathfrak{c}_3,\ \ldots, 2^{\mathfrak{c}_n} = \mathfrak{c}_{n+1}, \ldots$$

в прогрессии численностей

$$\mathfrak{a},\ \mathfrak{c},\ \mathfrak{c}_1,\ \mathfrak{c}_2,\ \ldots,\ \mathfrak{c}_n,\ \ldots$$

каждый член больше, чем предыдущий. Таким образом, данная прогрессия представляет собой пример нарастающей бесконечной прогрессии, состоящей из бесконечных численностей. Существует и такая численность, которая больше, чем любой член указанной прогрессии. Такую численность можно найти следующим образом. Пусть A представляет собой множество исчислимой численности, а C — непрерывной численности, множества $C_1, C_2, \ldots, C_n, \ldots$ — численностей $\mathfrak{c}_1, \mathfrak{c}_2, \ldots, \mathfrak{c}_n, \ldots$ соответственно. Путем соединения такого исчислимо бесконечного количества множеств, численность $\mathfrak{m}$ полученного таким образом множества M наверняка будет больше любого члена указанной выше прогрессии численностей. Правильность соотношения $\mathfrak{c}_n < \mathfrak{m}$ $(n = 1, 2, \ldots)$ вытекает из того, что $\mathfrak{c}_n < \mathfrak{c}_{n+1}$ и $\mathfrak{c}_{n+1}$ *не больше*, чем $\mathfrak{m}$ (так как $C_n \subseteq M$). А затем, исходя из численности $\mathfrak{m}$, нарастающая прогрессия образуется таким же образом, как она образовалась в предыдущем случае, исходя из значения $\mathfrak{a}$.

Также нетрудно убедиться в том, что если S представляет собой любое множество численностей, то существует такая численность $\mathfrak{n}$, которая больше, чем любой элемент множества S. Для построения такой численности $\mathfrak{n}$ для каждого элемента $\mathfrak{m}$ множества S выбирается соответствующее множество численности $\mathfrak{m}$. Эти множества соединяются. Ни один из элементов множества S не может быть больше, чем численность полученного таким образом множества, а численность степенного множества данного множества определенно будет больше, чем любой элемент множества S.

Остается открытым вопрос: если $\mathfrak{m}$ представляет собой *бесконечную* численность, то существует ли численность между численностями $\mathfrak{m}$ и $2^{\mathfrak{m}}$. Этот вопрос представляет собой обобщенную проблему непрерывности, из которой следует (в оригинальном смысле) рассмотренная проблема непрерывности в случае $\mathfrak{m} = \mathfrak{a}$.

УПОРЯДОЧЕННЫЕ МНОЖЕСТВА

УПОРЯДОЧЕНИЕ; ПОДОБИЕ; ТИПЫ ПОРЯДКОВ

Проблемы, оставшиеся открытыми в теории бесконечных множеств: трихотомия соотношения «$<$» между численностями и проблема непрерывности. Приближение к этим понятиям стало возможным в теории упорядоченных множеств.

Упорядоченное множество. Понятие упорядочения множества является обобщением понятия упорядочения чисел по их значению. Множество называется *упорядоченным*, если между его элементами определяется отношение порядка; для любой пары a, b элементов множества отношение «a опережает b» является действительным или ложным. Оно является *иррефлексивным, асимметричным, транзитивным и трихотомным.* В более подробной форме: если соотношение «a опережает b» обозначается символом «$a \prec b$», то для любых элементов a, b, c множества действительны следующие соотношения:

1. $a \prec a$ ложно	*(иррефлексивность)*,
2. если соотношение $a \prec b$ правильно, то соотношение $b \prec a$ является ложным	*(асимметрия)*,
3. если соотношения $a \prec b$ и $b \prec c$ являются правильными, то соотношение $a \prec c$ также правильно	*(транзитивность)*,
4. если выполняется условие $a \neq b$, то по крайней мере одно из соотношений $a \prec b$ и $b \prec a$ является правильным	*(трихотомия)*.

Множества естественных, рациональных и действительных чисел являются упорядоченными множествами относительно обычного соотношения «$<$». Но каждое из них является упорядоченным и относительно соотношения «$>$».

Множество называется *частично упорядоченным*, если между его элементами определяется такое отношение порядка, которое является иррефлексивным, асимметричным и транзитивным (но, возможно, не трихотомным). Например, между элементами степенного множества H соотношением $A \subset B$ устанавливается частичное упорядочение (см. гл. «Подмножество, действительная часть»); обычно здесь не выполняется условие трихотомии : если множество H имеет по крайней мере 2 элемента, то оно имеет такие части A и B, из которых ни одна не является частью другой

Подобие упорядоченных множеств. Два упорядоченных множества A и B называются *подобно упорядоченными*, (или *подобными)*, если существует такая обратимая функция f, областью существования которой является A, значениями функции — B. Если a_1 и a_2 являются элементами множества A, то соотношение $f(a_1) \prec f(a_2)$ является правильным тогда и только тогда (по упорядочению в множестве B), когда соотношение $a_1 \prec$ $\prec a_2$ является правильным (по упорядочению в множестве A). Таким образом, упорядоченные множества A и B являются подобными, если между их элементами может быть установлено предписание типа 1—1, держащее упорядочение. Функция f, осуществляющая такое упорядочение, называется *монотонной*. Символом подобия множеств A и B является «$A \simeq B$».

Из этого определения ясно, что только эквивалентные множества могут быть подобными ; далее, если множество A является упорядоченным и выполняется соотношение $A \sim B$, то множество B может быть упорядочено таким образом, чтобы выполнилось соотношение $A \simeq B$.

Пример: упорядочение множества натуральных чисел по соотношению «$<$» и упорядочение множества отрицательных целых чисел по соотношению «$>$» являются подобными ; с помощью обратимой и монотонной функции

$$f(n) = -(n+1) \qquad (n = 0, 1, 2, 3, \ldots)$$

первое упорядочение преобразуется во второе. Преобразование представлено в следующей таблице:

n:	0	<1	<2	<3	<4	$<5\ldots$	$<n<\ldots$
$f(n) = -(n+1)$	-1	>-2	>-3	>-4	>-5	$>-6\ldots$	$>-(n+1)>\ldots$

Замечание. Представленное здесь понятие *монотонной функции* содержит в себе и понятие *строго монотонной функции* (см. стр. 395). Если функция $f(x)$ является строго (монотонно) нарастающей на некотором интервале, то она преобразует этот интервал, упорядоченный по соотношению «$<$», (при соблюдении упорядочения) на множество чисел, упорядоченное также по соотношению «$<$». А с другой стороны, если функция $f(x)$ является строго (монотонно) убывающей на некотором интервале, то этот интервал, упорядоченный по соотношению «$<$», преобразуется данной функцией при соблюдении упорядочения на множество чисел, упорядоченное по соотношению «$>$».

Пустое множество и множества с одним элементом в любом случае рассматриваются как упорядоченные множества, так как условия для соотношений упорядочения не дают никакого ограничения в отношении этих множеств. Путем упорядочения множества из двух элементов получится упорядоченная пара (транзитивность и в этом случае не имеет значения). Можно убедиться (путем применения полной индукции) в том, что при $n > 2$ любое упорядочение конечного множества с n элементами подобно упорядочению естественных чисел $(1, 2, 3, \ldots, n-1, n)$ по величине.

Легко продумать, что подобие упорядоченных множеств как и эквивалентность множеств) является рефлексивным, симметричным и транзитивным соотношением. (См. стр. 553).

Типы порядков. Для охарактеризования подобно упорядоченных множеств введено понятие *типа порядков.* Каждому упорядоченному множеству предписывается отдельный знак, который носит название *типа порядка* множества с тем, что типы порядков подобно упорядоченных множеств совпадают, а типы порядков неподобно упорядоченных множеств отличаются. Тип порядка множества A обозначается символом $\bar{A}$. Для самостоятельного обозначения типов порядков применяются греческие строчные буквы.

Под определением численности типа порядка α подразумевается численность такого упорядоченного множества A, для которого $\bar{A} = \alpha$ (т. е. численность некоторого множества типа порядка α).

Поскольку множества одинаковых типов порядка одновременно имеют и одинаковые численности, безразлично, численностью какого конкретного множества (типа порядка α) задается численность типа порядка α. Численность типа порядка α обозначается симводом $\bar{\alpha}$. Таким образом, если $\bar{A} = \alpha$, то $\bar{\alpha} = \bar{\bar{A}}$.

Так как любые два упорядочения некоторого конечного множества являются подобными, а упорядочение двух множеств различных численностей никогда не может быть подобным, то типы порядков конечных множеств однозначно определяются количеством их элементов. Именно поэтому тип порядка некоторого конечного множества обозначается таким же образом, как и его численность.

Тип порядка множества естественных чисел, упорядоченного по их величине (т. е., по соотношению «$<$»), обозначается через ω. Ясно, что каждое множество, упорядоченное в прогрессию и представляющее собой исчислимое множество, является множеством типа порядка ω.

Упорядоченное по соотношению «$>$» множество естественных чисел, т. е. обратное упорядочение

$$\ldots \prec 5 \prec 4 \prec 3 \prec 2 \prec 1 \prec 0$$

обозначается символом ω^* типа порядка. Таким же типом порядка обозначается упорядочение множества отрицательных целых чисел по соотношению «$<$»;

$$\ldots < -5 < -4 < -3 < -2 < -1.$$

Типы порядка множеств рациональных и действительных чисел, упорядоченных по их величине, обозначаются через η и λ соответственно. Любой открытый интервал действительных чисел, упорядоченных по их величине, также имеет тип порядка λ.

В области типов порядка могут быть определены также и операции (эти вопросы здесь рассматриваться не будут).

ДОЛЯ УПОРЯДОЧЕННОГО МНОЖЕСТВА

Под определением доли упорядоченного множества A подразумевается такая часть A' множества A, которая — вместе с любым элементом — содержит в себе каждый *предыдущий* элемент. (Элемент a является предыдущим элементом элемента b, если выполняется соотношение $a \prec b$ по данному упорядочению.)

Доля A' является *действительной* долей множества A, если, с одной стороны, A' является долей, а с другой стороны — действительной частью множества A (т. е. $A' \neq A$). Любое упорядоченное множество является всегда (недействительной) долей самого себя.

Пример: Множество целых чисел, каждое из которых меньше 1000, является действительной долей упорядоченного по величине множества целых чисел, а множество рациональных чисел, каждое из которых меньше, чем $\sqrt{2}$, — действительной долей упорядоченного по величине множества рациональных чисел.

Доли, составленные по *Дедекинду*, служащие моделью теории множеств действительных чисел, на самом деле представляют собой некоторые частные доли упорядоченного по величине множества рациональных чисел. Множество тех элементов некоторого упорядоченнного множества, которые предшествуют элементу a множества A, представляет собой действительную долю множества A. Такая доля называется долей, *генерированной элементом* a, и обозначается символом $A|a$.

В предыдущих двух примерах первый пример представляет собой долю, «генерированную элементом» (генерирующий элемент — 1000), а второй пример — нет ($\sqrt{2}$ не представляет собой рациональное число и можно доказать, что вышеприведенная доля не может быть генерирована никаким рациональным числом).

ХОРОШО УПОРЯДОЧЕННЫЕ МНОЖЕСТВА; ПОРЯДКОВЫЕ ЧИСЛА

Конечными типами порядка характеризуются конечные прогрессии, а типом порядка ω обозначается упорядочение в бесконечную прогрессию. Множество, численность которого больше чем исчислима, не может быть упорядочено в прогрессию. В то же время может быть представлено такое обобщение понятия прогрессии, с помощью которого несколько прогрессий могут быть соединены друг за другом. Например, в таблице на стр. 556 множество положительных дробей собрано в исчислимо бесконечном количестве прогрессий, а соединение друг за другом этих прогрессий дает в результате такое упорядочение, которое может быть названо *трансфинитной прогрессией*. Нетрудно представить себе, что путем соединения больше чем исчислимо бесконечного количества прогрессий и множества, больше, чем исчислимой численности, могли бы быть упорядочены в трансфинитной прогрессии, и таким образом их «длина» могла бы быть сравнена, и в этом случае проблемы трихотомии и непрерывности могли бы быть решены таким путем.

Таким образом, из наглядного понятия «трансфинитной прогрессии» удалось создать четкий математический термин путем введения понятия «хорошо упорядоченного множества».

ХОРОШО УПОРЯДОЧЕННЫЕ МНОЖЕСТВА

Упорядоченные множества называются *хорошо упорядоченными*, если в них любая пустая часть имеет «первый» элемент, т. е. такой элемент, которому по данному упорядочению не предшествует ни один элемент данного подмножества.

Могут быть доказаны следующие установления:

1. Некоторое упорядоченное множество является хорошо упорядоченным тогда и только тогда, если оно не имеет части типа ω^*.

2. Отсюда следует и можно доказать, что любое упорядочение конечного множества является хорошим упорядочением.

3. Любая действительная доля хорошо упорядоченного множества всегда является долей, генерированной некоторым ее элементом.

4. Каждый элемент хорошо упорядоченного множества, за исключением «последнего» его элемента (если оно имеет такой элемент) имеет «*непосредственно последующий элемент*» (*элемент b является непосред-*

ственно последующим элементом элемента a, если $a \prec b$ и не существует такой элемент c, для которого выполняются соотношения $a \prec c$, $c \prec b$).

5. Каждая часть хорошо упорядоченного множества является также хорошо упорядоченной (по оригинальному упорядочению).

6. Если A и B представляют собой хорошо упорядоченные и подобные множества, то существует только одна такая функция f, с помощью которой множество A преобразуется в множество B при соблюдении упорядочения (в противовес, например, тому, что два эквивалентных множества могут быть предписаны друг другу по-разному, или, что иногда между двумя подобными, но не хорошо упорядоченными множествами могут быть определены разные преобразования при соблюдении упорядочения).

7. Некоторое хорошо упорядоченное множество никогда не может быть подобным своей действительной доле.

8. Если два хорошо упорядоченных множества не подобны, то одно из них (но только одно) обязательно подобно некоторой действительной доле другого множества.

Упорядоченное по величине множество натуральных чисел представляет собой хорошо упорядоченное множество, так как в таком упорядочении натуральному числу n предшествуют только числа $0, 1, 2, \ldots, n-1$, и, исходя из его единственного элемента, нельзя составить бесконечную, обратно идущую прогрессию, т. е. образовать часть типа ω^* (см. теорему 1). Упорядоченное по величине множество рациональных и действительных чисел не является хорошо упорядоченным (для любого рационального и действительного числа может быть найдено бесконечное число стоящих ниже по значению рациональных и действительных чисел). Некоторые множества можно по-разному не только упорядочить, но и хорошо упорядочить. Например, такое упорядочение множества натуральных чисел, при котором сначала идут четные числа, а затем нечетные числа (по значению):

$$0 \prec 2 \prec 4 \prec 6 \prec \ldots \prec 1 \prec 3 \prec 5 \prec 7 \prec \ldots$$

представляет собой также хорошее упорядочение и не подобно упорядочению по величине.

Естественно, что каждое исчислимое множество может быть хорошо упорядочено, так как такое множество всегда может быть упорядочено в прогрессию и каждая прогрессия подобна упорядочению типа ω множества натуральных чисел.

Из теорем 1—8 видно, что хорошее упорядочение во многом напоминает упорядочение в прогрессию. Далее будет показано, что к элементам некоторого хорошо упорядоченного множества могут быть предписаны и *индексы* (как к членам обычных прогрессий). Для этого необходимо такое распространение понятия натуральных чисел, на основе которого интерпретируется достаточное количество индексов для элементов множества любой большой численности. С такой целью создано понятие *порядкового числа*.

ПОРЯДКОВЫЕ ЧИСЛА

Типы порядка хорошо упорядоченных множеств называются *порядковыми числами*. Поскольку любое упорядочение конечного множества является хорошим упорядочением, то конечные типы порядков, т. е. все естественные типы порядков являются порядковыми числами.

В области порядковых чисел соотношение «меньше» определяется следующим образом. *Известно, что α меньше β ($\alpha < \beta$), если хорошо* упорядоченное множество типа порядка β имеет действительную долю типа порядка α. Однозначность данного определения доказывается на основе применения понятий подобия, хорошего упорядочения и теоремы 6. Определенное таким образом соотношение «$<$» является иррефлексивным, асимметричным, транзитивным и трихотомным. Иррефлексивность этого соотношения обеспечивается теоремой 7. Транзитивность получается непосредственно путем применения теоремы 6. Асимметрия в любом случае следует из иррефлексивности и транзитивности. Трихотомия является следствием теоремы 8.

Но из теоремы 8 также следует, что из двух хорошо упорядоченных множеств по крайней мере одно является эквивалентным части другого. Ведь если множества A и B являются хорошо упорядоченными и подобными, то $A \sim B$ (тогда каждое множесто эквивалентно части другого), а если они не подобны, то на основе теоремы 8 одно из них подобно действительной доле другого множества. Но поскольку доля представляет собой также часть и эквивалентность включает в себя подобие, то в данном случае одно из множеств эквивалентно части другого. Таким образом, среди хорошо упорядоченных множеств нет таких, которых нельзя сравнивать между собой : если два таких множества не являются эквивалентными, то одно из них эквивалентно части другого. Итак, для численностей хорошо упорядоченных множеств действительно соотношение трихотомии. Если каждое множество является хорошо упорядоченным, то трихотомия относительно численностей также правильна.

Под *индексом* некоторого элемента a хорошо упорядоченного множества A подразумевается тип порядка (это всегда порядковое число) множества (по оригинальному упорядочению) предшествующих элементов, то есть доли $A|a$.

Если индексом некоторого элемента является α, то элемент может быть обозначен a_α.

Легко убедиться в том, что в хорошо упорядоченном множестве соотношение $a_\alpha \prec a_\beta$ является правильным тогда и только тогда, когда $\alpha < \beta$. Доля, генерированная первым элементом хорошо упорядоченного множества, представляет собой пустое множество, поэтому индексом первого элемента является 0. Индекс ω имеет тот элемент, которому предшествует обычная бесконечная прогрессия, то есть упорядоченное множество типа ω (как единице в примере на стр. 571). Можно показать, что любое множество порядковых чисел является хорошо упорядоченным относительно соотношения «$<$». Некоторое множество порядковых чисел называется *беспромежуточным* в том случае, если вместе с любым порядковым числом данного множества содержит и каждое меньшее порядковое число ; элементы такого множества R будут те и только те порядковые числа, которые меньше, чем тип порядка R.

Трансфинитные прогрессии. Аналогично тому, что функции, определенные в множестве естественных чисел, называются прогрессиями, функции, определенные в множестве беспромежуточных порядковых чисел, называются *трансфинитными прогрессиями*. (Настоящее определение включает в себя и понятие обычной бесконечной прогрессии, и конечной прогрессии.)

Можно доказать, что множество порядковых чисел, примененных для индексации элементов некоторого хорошо упорядоченного множества, является всегда беспромежуточным и подобным данному хорошо упорядоченному множеству. Таким образом, элементы с индексами некоторого хорошо упорядоченного множества всегда представляют собой трансфинитную прогрессию. Обратное также верно: пусть R представляет собой множество беспромежуточных порядковых чисел, A — значения обратимой функции f, определенной в множестве R, и при выполнении соотношения $\alpha \in R$ значение функции $f(\alpha)$ обозначается через a_α. В таком случае хорошее упорядочение множества A определяется следующим образом. Условие $a_\alpha < a_\beta$ выполняется тогда и только тогда, если $\alpha < \beta$ (α и β здесь являются элементами множества R). Тогда выполняется соотношение $R \cong A$, и поскольку R является хорошо упорядоченным (как множество порядковых чисел), то множество A стало также хорошо упорядоченным. Итак, понятие хорошего упорядочения аналогично упорядочению его элементов в трансфинитную прогрессию. В связи с численностями хорошо упорядоченных бесконечных множеств можно показать, что они могут быть упорядочены аналогично порядковым числам. Для их обозначения Кантором применялась первая буква еврейского алфавита, алеф ($\aleph$). Алефы, дополненные индексами порядковых чисел, являются численностями хорошо упорядоченных бесконечных множеств. Таким образом $\aleph_0 = \mathbf{a}$ является наименьшим алефом и $\aleph_\alpha < \aleph_\beta$ тогда и только тогда, когда $\alpha < \beta$.

Теорема о хорошем упорядочении. Впервые Цермело доказал (1908 г.), что предположение, по которому каждое множество является хорошо упорядочимым, равноценно его основному тезису:

Если элементы некоторого множества A являются попарно дизъюнктивными, то существует такое множество B, у которого имеются именно по одному общему элементу с каждым элементом множества A.

Итак, множество B содержит в себе по «одной пробе» и по одному «избранному» элементу каждого элемента множества A (нельзя забывать о том, что элементы множества A также множества!)

Эту теорему не удалось упростить, и она принята как аксиома в области теории множеств (как теорема, не противоречающая обычному понятию наглядности при использовании ее для конечных множеств) и называется *аксиомой выдора*. Путем применения аксиомы выбора может быть доказана так называемая *теорема о хорошем упорядочении*, по которой каждое множество является хорошо упорядочимым. (Как упоминалось выше, вследствие настоящей теоремы, в области численностей понятие *трихотомии* имеет общее значение).

На основе теоремы о хорошем упорядочении, каждая бесконечная численность равняется некоторому алефу. Таким образом, действительно соотношение $\mathbf{c} = \aleph_\alpha$ для некоторого значения α (случай $\alpha = 0$ исключается) и представлены такие результаты, которые показывают, с какими порядковыми номерами нельзя считаться. После доказательства теоремы о хорошем упорядочении проблема непрерывности снова осталась нерешенной.

ПРОБЛЕМЫ ТЕОРИИ МНОЖЕСТВ

МАТЕМАТИКА И ТЕОРИЯ МНОЖЕСТВ

Значение теории множеств для других глав математики прежде всего заключается в том, что в большей или меньшей степени ее методы и результаты применяются почти во всех главах математики. Примеры таких применений представлены и в таких главах настоящей книги, как «Анализ», «Алгебра», «Геометрия».

Теория множеств позволяет создать *модель* любого известного раздела математики и, таким образом, теория множеств объединяет эти разделы в единую систему.

Под термином «модели» по теории множеств одной из частей математики подразумевается следующее. Из понятий данной главы выбираются несколько таких, с помощью которых другие понятия системы могут быть определены через логические определения. Эти выбранные понятия заменяются «соответственно выбранными» понятиями теории множеств. Выражение «соответственно выбранное» означает, что если заменить понятия, присутствующие в некоторой теореме первоначальной системы соответствующими понятиями теории множеств, то «перевод» теоремы, полученный таким образом, представляет собой правильную теорему в теории множеств. (Если теоремы первоначальной системы приведены из некоторых основных теорем (так называемых аксиом), путем применения логических доказательств, то достаточно обеспечить, чтобы «переводы» аксиом представляли собой правильные теоремы в теории множеств.) Система понятий в теории множеств, заменяющих основные понятия первоначальной системы, считается моделью первоначальной системы понятий.

Например, рассмотрим модель по теории множеств арифметики естественных чисел. Показано, что операции, проводимые на естественных числах, могут быть выражены через понятия пересчета на 1. Таким образом, достаточно заменить эти два понятия понятием из теории множеств: эти два понятия представляют собой понятия пересчета на 0 и 1. Заменой 0 выбирается пустое множество ($\emptyset$). Операция пересчета заменяется такой операцией из теории множеств, с помощью которой из множества A образуется множество $A \cup \{A\}$. Таким образом, если $A = \emptyset$, тогда

$$A \cup \{A\} = \emptyset \cup \{\emptyset\} = \emptyset.$$

Путем применения пересчета на 0 образуется 1. Таким образом, замена 1 по теории множеств получится, если к замене 0, $\emptyset$ применяется операция, заменяющая пересчет. Это приводит к множеству $\{\emptyset\}$. Если еще раз применить вышеуказанную операцию, получится замена 2 по теории множеств:

$$A \cup \{A\} = \{\emptyset\} \cup \{\{\emptyset\}\} = \{\emptyset, \{\emptyset\}\}.$$

Это множество, состоящее из двух элементов, представляет собой замену 2. Заменой 3 является следующее множество, состоящее из трех элементов и образуемое аналогичным образом:

$$\{\emptyset, \{\emptyset\}, \{\emptyset, \{\emptyset\}\}\},$$

и так далее.

В отношении эффективности математических моделей можно отметить, что если в «надежной» теории удастся создать модель другой теории, то другая теория также будет надежной.

ПРОТИВОРЕЧИЯ В ТЕОРИИ МНОЖЕСТВ

В начале XX века оказалось, что *наивная* теория множеств представляет собой *противоречивую* теорию. Математическая теория считается противоречивой, если в ней имеется такая доказуемая теорема, отрицание которой также можно доказать. Ниже представлено несколько примеров противоречий в теории наивной теории множеств:

Противоречие множества всех предметов. Для любого множества H действительно следующее соотношение:

(а) $$\overline{\overline{H}} < \overline{\overline{2^H}}$$

(см. стр. 565). Но если множество H представляет собой множество всех предметов, то $2^H \subseteq H$ (так как *все предметы* являются элементами множества H), и поэтому 2^H *не больше, чем* H. В то же время это соотношение представляет собой отрицание соотношения (а).

Противоречие Расселя. Пусть некоторое множество называется «порядочным», если оно не является элементом самого себя. Пусть A — множество всех порядочных множеств. Тогда

(б) *Множество A является или порядочным, или непорядочным.*

Если бы множество A было порядочным, то оно являлось бы элементом самого себя (так как каждое порядочное множество является элементом множества A), то есть, не могло бы быть порядочным. Если бы оно не было порядочным, то есть оно было бы элементом самого себя, то все-таки было бы порядочным (так как каждый элемент множества A является порядочным). Таким образом, *исключено и то, что множество A является порядочным, и то, что оно является непорядочным.* Но это отрицание соотношения (б).

Противоречие Ричарда. Проанализируем следующий текст:

> Пусть даны все тексты величиной по крайней мере 500 знаков с помощью фиксированной пишущей машинки (пропуск между словами также является знаком). Пусть x — наименьшее естественное число, определения которого в русском языке нельзя найти среди этих текстов.

Во-первых, нетрудно убедиться, что количество текстов, определенных первым предложением, является ограниченным (если, скажем, имеется 100 знаков на пищущей машинке, то количество таких текстов составляет 100^{500}). Таким образом, множество таких естественных чисел, которые определяются подобными текстами на русском языке, является ограниченным. Следовательно, определенное по данному тексту число x существует, и по его определению число x *не может быть определено по крайней мере 500 знаками* (в соответствии с указаниями данного текста). В то же время данный текст во всех отношениях соответствует своим предписаниям и определяет именно число x. Поэтому *число x все-таки может быть определено по крайней мере 500 знаками*, в соответствии с предписанием. Наша система логических выводов является такой системой, в которой при одном противоречии любое соотношение, как и отрицание любого соотношения, может быть доказано (см. гл. «Математическая логика»). Поэтому логическое противоречие не допускается в некоторой математической системе.

Присутствие противоречий показывает, что следует ограничить такую точку зрения наивной теории множеств, по которой любые предметы могут быть элементами некоторого множества. В отношении вида ограничений возникли различные предложения. По мнению представителей так называемого направления *интуиционистов*, следует полностью отказаться от множеств, численность которых более, чем исчислимая. Такое суждение ведет к существенным искажениям математики. По оригинальной программе *логицистов* теория множеств рассматривается как часть логики; во избежание противоречий ими используется следующий прием: все объекты, которые присутствуют на обеих сторонах соотношения» «$\in$», разделяются на типы, а типы упорядочены в некоторую иерархию. Затем устанавливается, что предметы, присутствующие на правой стороне соотношения «$\in$», должны принадлежать типу на 1 больше, чем предметы на левой стороне. Так как данное ограничение привело к определенным затруднениям при создании теории, ограничение было частично снято применением так называемой аксиомы сократимости.

Наконец, по третьему направлению — теории *Цермело* — понятие множества ограничивается таким образом, что задаются некоторые начальные множества и несколько операций, и лишь те множества считаются предметом теории множеств, которые возникают из начальных множеств с помощью заданных операций.

Общей чертой последних двух направлений является использование *аксиоматического метода*. Аксиоматический метод применяется не только в теории множеств, но и во всех главах современной математики, а также и в других областях наук, пользующихся математическими методами.

АКСИОМАТИЧЕСКАЯ ТЕОРИЯ МНОЖЕСТВ

Аксиоматическое построение некоторой теории (отрасли науки) имеет следующий вид:

а) Задаются некоторые понятия теории как неопределенные *основные понятия*.

б) Без доказательства задаются некоторые соотношения, относящиеся к основным понятиям. Они называются *аксиомами*. В определениях аксиом могут присутствовать только основные и логические понятия. Совокупность основных понятий и аксиом составляет *систему аксиом*.

в) Следующие понятия теории определяются с помощью логических определений, из основных понятий. Они представляют собой *определенные понятия*. В их определениях могут присутствовать только основные понятия, уже определенные понятия и логические понятия.

г) При более строгом аксиоматическом построении, кроме основных понятий, представляются логические понятия, а наряду с аксиомами — правила применяемых логических выводов.

Ниже кратко излагается так называемая *система аксиом* теории множеств по *Цермело и Френкелю*.

Основные понятия: понятия *множества* и *принадлежности к множеству* (принадлежность к элементу множества). Определенные понятия (например, подмножество, степенное множество) получаются из них таким образом, как и в наивной теории множеств. При формулировании аксиом (с целью краткого изложения) мы здесь применяем и определенные понятия). (Отметим, что в этой системе учитываются только множества как элементы множеств. Это особенно важно для того, чтобы понять аксиомы 5 и 9.)

Аксиомы

1. Если элементы двух множеств одинаковы, то эти два множества тождественны между собой.

2. Существует такое множество, которое не имеет ни одного элемента (пустое множество, $\emptyset$).

3. Если a и b представляют собой множества, тогда существует такое множество, элементами которого являются именно a и b (это будет множество $\{a, b\}$; или в случае выполнения соотношения $a = b$ — множество $\{a\}$).

4. Существует такое множество, элементом которого является $\emptyset$, и если множество a является его элементом, то $\{a\}$ также является его элементом (существование бесконечного множества).

5. Для любого множества существует такое множество, которое является соединением элементов данного множества.

6. Для любого множества существует его степенное множество.

7. Если A является множеством, а T — такой характерной чертой, которая может быть определена с помощью понятий теории множеств, то существует множество таких элементов множества A, которые имеют характерную черту T.

8. Если A является множеством и f — такой функцией, которая может быть определена с помощью понятий теории множеств, область ее существования — A, а ее значения являются множествами, то значения функции f также представляют собой множество.

9. Непустое множество всегда имеет отличный от него элемент.

10. Аксиома выбора (см. стр. 573).

В теории множеств хорошо известны системы аксиом Неймана и Бернайса; обе системы равноценны системам *Цермело и Френкеля*. Противоречия в них (до сих пор) не установлены, хотя этому нет достоверных доказательств. В системе аксиом теории множеств по логицистам применяется существенно меньшая часть наивной теории множеств.

В системе аксиом Неймана в сущности применяются не термины «множества» и «его элемента», а понятия «функции» и «ее аргумента»; множества выражаются в ней через их характеристические функции (см. стр. 551).

В системе аксиом Бернайса кроме понятий «множества» и «его элемента», как основное понятие фигурирует понятие «класса». Классы являются *размерами характерных черт* (под термином размера некоторой характерной черты подразумеваются такие предметы, которые имеют данную характерную черту). Для любого множества A существует такой класс, элементы которого совпадают с элементами множества A, что является размером характерной черты (быть элементом множества A). В то же время не каждому классу принадлежит такое множество, элементы которого совпадают с элементами данного класса. Например, размеры характерных черт «быть множеством», «быть численностью», «быть порядковым номером» являются примерами таких классов. В такой системе можно говорить *о классе* всех множеств, но нельзя говорить *о множестве* всех множеств. Элементами классов как и множеств могут быть исключительно множества. (Как мы уже убедились, понятие множества всех множеств является противоречивым; оно не может присутствовать в аксиоматической теории. Решение Бернайса позволяет нам говорить о всех множествах без ссылки на противоречивые понятия. В системе Неймана также присутствует фактически ровноценный этому прием.)

В теории множеств проблема выборочной аксиомы долгое время являлась спорной. Многие считали, что хотя установление аксиомы для конечных множеств не противоречит наглядности, но оно может стать источником противоречий для бесконечных множеств. К. Гёделем (1938) было доказано, что выборочной аксиомой не вносится противоречие в системы известных аксиом теории множества: если такие системы являются непротиворечивыми *без* выборочной аксиомы, то и при наличии выборочной аксиомы в них не будет противоречий.

В связи с проблемой непрерывности также Гёделем было доказано, что предположение

$$2^{\aleph_\alpha} = \aleph_{\alpha+1}$$

не приводит к противоречию в известных системах аксиом теории множества.

П. Коханом (1963) было доказано, что и *отрицание* данного предположения может быть согласовано с другими аксиомами. Таким образом, проблема непрерывности в известных системах аксиом теории множеств является *неразрешимой* проблемой. Это означает, что вышеуказанное соотношение, либо его отрицание, может быть присоединено к любой известной системе аксиом как новая аксиома, либо можно поставить эксперимент для создания новой системы аксиом, отличающейся от известных.

КРАТКИЙ ОБЗОР ИСТОРИИ ТЕОРИИ МНОЖЕСТВ

Одним из методов математики является метод применения абстракции, при которой не принимаются во внимание некоторые конкретные обстоятельства. Это неизбежно приводит к возникновению понятия множества, которое стало основным понятием математики.

Теория множеств как самостоятельная дисциплина сформировалась сравнительно поздно, в конце XIX века. Основателем ее был немецкий математик Р. Дедекинд (1831—1916), в работах которого впервые начали формироваться признаки понятия множества.

Действительным основоположником теории множеств считается Г. Кантор (1845—1918) — немецкий математик русского происхождения, исследования которого по теории множеств были основаны на приведении не непрерывных функций на основе ряда Фурье. Работы Кантора были настолько абстрактными, что это приводило к острым спорам во время публикаций, и только позже стали общепризнанными.

В развитие теории множеств значительный вклад внесли следующие ученые мира: Н. Н. Лузин (1883—1950), М. Я. Суслин (1894—1919), П. С. Александров (1896—) и А. Н. Колмогоров (1903—) советские, П. Бернайс (1888—) — швейцарский, Э. Цермело (1871—1953), Ф. Бернштейн, Ф. Хаусдорф, Г. Хессенберг, А. Шёнфлис — немецкие, Бендиксон — шведский, К. Куратовский (1896—), В. Шерпинский (1882—1969) польские, Иоганн фон Нейман (1903—1957) американский (венгерского происхождения), Дьюла Кёниг (1849—1913) и Денеш Кёниг, Л. Калмар (1905—), Р. Петер (1905—) и Я. Шурани (1918—) — венгерские.

ТЕОРИЯ ВЕРОЯТНОСТИ

ЗАДАЧА ТЕОРИИ ВЕРОЯТНОСТИ

Опыт показывает, что среди случайных явлений существуют определенные закономерности, но только в том случае, когда наблюдается *большое число случайных явлений одновременно*.

Задача теории вероятности: *математическое исследование закономерностей массовых случайных явлений*.

ЧАСТОТА

Пусть проводятся n экспериментов при одинаковых обстоятельствах. Пусть результатом каждого эксперимента будет наступление или ненаступление некоторого события A.

Примеры: 1. Пусть событие A означает, что монета выпадает на решку и монета подбрасывается n раз.

2. Пусть событие A означает, что в девятом классе самый высокий ученик выше, чем 170 см, и тогда рассматривается n случайно выбранных девятых классов.

3. Пусть событие A означает, что 1 марта уровень воды Дуная в Будапеште выше, чем 500 см, и пусть уровень воды Дуная рассматривается 1 марта в n различных годах.

Предположим, что из n экспериментов событие наступило k раз и не наступило $n-k$ раз. Тогда число k называется частотой события A, а частное $\frac{k}{n}$ — относительной частотой в данной последовательности экспериментов.

Наш опыт показывает, что при увеличении количества экспериментов *относительная частота некоторого события* — при выполнении определенных условий — все более и более приближается к некоторому числу, которое называется *вероятностью* события A.

Эксперименты могут быть следующие: из некоторого объема товаров случайным образом (например, по жребию) выбираются несколько штук и рассматриваются выбранные товары (такой метод часто применяется при контроле качества), или задаются вопросы некоторым случайно выбранным людям, чтобы узнать их мнение о некоторых вопросах. Например, для торговли весьма существенно знать, какой ожидается спрос на некоторые товары.

Такой вопрос исследуется группами так называемых исследователей рынка сбыта таким образом, что они задают вопросы случайно избранным людям для выяснения намерения их в отношении покупки и размера покупок данного товара. На основе собранных таким образом ответов они делают выводы относительно потребностей всего населения.

Конечно, не безразлично, каким образом были выбраны лица, которым были заданы вопросы. Например, на некоторые товары может оказаться большим спрос в городах, чем в деревнях; спрос может быть различным у интеллигенции и у рабочих, у пожилых и у молодых людей и т. д. Это означает, что при разработке плана выбора людей, которым задаются вопросы, следует учитывать эти различия.

В некоторых случаях вероятность некоторого события может быть определена и без проведения экспериментов. *Задача теории вероятности заключается в разработке математических методов для расчета вероятностей некоторых событий.* Методы теории вероятности приемлемы только в том случае, когда известны вероятности некоторых *более простых* событий, чтоды исходя из этих вероятностей, можно было рассчитывать вероятности *более сложных* событий. Например, на основе математических методов (по теории вероятности) нельзя дать ответ на вопрос, в самом ли деле монета падает решкой с вероятностью $\frac{1}{2}$. Это зависит от того, является ли монета симметричной, не тяжелее ли одна ее сторона, не изношена ли некоторая ее часть и т. д. Но если нам известно, что монета падает решкой действительно с вероятностью $\frac{1}{2}$, то на основе методов теории вероятности можно ответить на вопросы: какая вероятность того, что монета дважды подряд падает решкой, или какая вероятность того, что после 100 бросков относительная частота падания орлом будет между границами 0,45 и 0,55.

При известных вероятностях некоторых простых событий наша цель заключается в разработке расчетов вероятностей других, более сложных событий. Математическим обсуждением связей между событиями занимается алгебра событий.

АЛГЕБРА СОБЫТИЙ

ОПЕРАЦИИ НА СОБЫТИЯХ

СУММА СОБЫТИЙ

Под суммой событий A и B $(A + B)$, подразумевается такое событие, при котором из событий A и B наступит хотя бы одно.

Пример: Пусть A означает такое событие, что при подбрасывании двух монет обе падают на одну и ту же сторону. Пусть B означает, что первая монета падает решкой. В таком случае $A + B$ означает, что результатом подбрасывания является или ОО (орел, орел), или РР, или РО, то есть $A + B$ означает, что результат подбрасывания не будет ОР.

Нетрудно убедиться в том, что при любых событиях A и B правильны следующие соотношения: $A + A = A$, $A + B = B + A$ и $A + (B + C) = (A + B) + C$.

ПРОИЗВЕДЕНИЕ СОБЫТИЙ

Объяснение понятию произведения событий можно дать подобно тому, как это выполняется при суммировании событий.

Пусть A и B представляют собой два события. Под определением произведения событий A и B (AB) подразумевается такое событие, что оба события A и B произойдут.

Пример: Пусть A означает, что на игральной кости выпадает четное число, а B — что на игральной кости выпадает число больше, чем 3, тогда AB означает, что на игральной кости выпадает 4 или 6.

Легко убедиться в том, что для любых событий A и B правильны следующие соотношения:

$$\boxed{AA = A, \quad AB = BA \quad \text{и} \quad A(BC) = (AB)C\,.}$$

ОБЯЗАТЕЛЬНОЕ И НЕОБЯЗАТЕЛЬНОЕ СОБЫТИЯ

Событие, которое наступит наверняка, называется обязательным событием.

Обязательным событием является, например, то событие, при котором игральная кость падает на 1, на 2, на 3, на 4, на 5 или на 6. Обязательное событие в дальнейшем обозначается через Ω.

Посмотрим, какие получатся результаты, если проводить операции сложения и произведения на обязательное событие.

Если прибавить любое событие A к обязательному событию Ω, то в результате получится обязательное событие, так как наверняка произойдет одно из событий A и обязательное событие. Таким образом:

$$\boxed{A + \Omega = \Omega\,.}$$

Аналогичным образом произведение некоторого события A на обязательное событие Ω представляет собой событие A, так как то, что оба события A и Ω произойдут, обозначает, что произойдет событие A. Итак,

$$\boxed{A\Omega = A\,.}$$

Аналогично термину обязательное событие вводится понятие *невозможного события.*

Невозможным называется такое событие, которое обязательно не произойдет.

Например, невозможным является событие AB, если A означает, что на игральной кости выпадает четное число и B означает, что результатом подбрасывания является 3. Невозможное событие обозначается буквой Φ.

Снова рассмотрим правила операции. Подобно предыдущему, можно убедиться в том, что при любом событии A:

$$\boxed{\Phi + A = A \qquad \Phi A = \Phi.}$$

РАЗНОСТЬ СОБЫТИЙ

Пусть A и B представляют собой два события. Под определением разности $A-B$ события A и события B подразумевается такое событие, при котором событие A произойдет, а событие B — нет.

Пример: Пусть A означает, что на игральной кости выпадет четное число ; B — больше, чем 2. В таком случае $A-B$ означает, что на игральной кости выпадает четное число, но не больше, чем 3, т. е. 2.

Посмотрим, что представляет собой событие $\Omega-A$. Это означает, что Ω произойдет, а событие A — нет ; поскольку событие Ω произойдет обязательно, поэтому $\Omega - A$ равняется тому, что событие A не произойдет. Такое событие обозначается символом $\bar{A}$ и называется *противоположностью* события A. Итак:

$$\boxed{\Omega - A = \bar{A}.}$$

Пример: Пусть A обозначает, что игральная кость выпадает на четное число. Тогда событие $\bar{A}$ означает, что игральная кость выпадает на нечетное число.

Легко увидеть, что при любом событии A

$$\boxed{A + \bar{A} = \Omega, \qquad A\bar{A} = \Phi.}$$

При рассмотрении приведенных выше правил операций заметно, что события создают булеву алгебру. Единичным элементом данной булевой алгебры является обязательное событие, а нулевым элементом — невозможное событие.

ТЕОРИЯ КОЛМОГОРОВА

ПРОСТРАНСТВО СОБЫТИЙ, ЭЛЕМЕНТАРНОЕ СОБЫТИЕ, СОБЫТИЕ

Предположим, что проводится экиперимент, результат которого зависит от случайности.

Например: 1. Подбрасывается монета (возможные результаты: орел, решка).

2. Подбрасывается игральная кость (возможные результаты: 1, 2, 3, 4, 5, 6).

3. Проводится измерение температуры воздуха; возможные колебания темпратуры на территории Венгрии: от -30 °C до $+40$ °C.

4. С помощью сейсмографа в течение суток проводятся измерения подземных толчков. Результатом измерения является функция (сейсмограмма).

Возможные результаты экспериментов всегда можно считать элементами некоторого множества.

Множество, содержащее все возможные результаты некоторого эксперимента, называется пространством событий и обычно обозначается символом Ω. Точки пространства событий, то есть возможные результаты эксперимента, называются элементарными событиями.

В первом примере множество событий Ω содержит в себе 2 элемента: орел, решка;. во втором примере множество событий Ω содержит в себе 6 элементов: 1, 2, 3, 4, 5, 6; в третьем примере множество событий Ω содержит в себе все возможные величины измеряемой температуры в области от -30 до $+40$ °C; в четвертом примере Ω содержит в себе все такие функции, независимая переменная которых изменяется от 0 до 24 часов, и которые могут возникать как сейсмограммы.

Событие A, связанное с некоторым экспериментом, может быть определено таким образом: при наступлении каких элементарных событий произойдет данное событие A. Иными словами, любому событию A может соответствовать некоторое подмножество пространства событий Ω, а именно множество тех элементарных событий, при наступлении которых произойдет событие A.

Например: если A означает, что подброшенная игральная кость выпадет на четное число, то событие A соответствует подмножествам пространства событий Ω, содержащим элементы 2, 4, 6. Если A означает, что в данный момент температура на территории Венгрии ниже 0°C, то событию A соответствует такое подмножество пространства событий Ω, которое содержит величины измеряемых температур между границами от -30 до 0°C.

Заметим, что при таком подходе естественно, что события образуют булеву алгебру, так как известно, что подмножества некоторого множества образуют булеву алгебру. Очевидно, что при таком подходе операции сложения событий соотвествует соединение множеств, произведению — образование общей части по теории множеств, а вычету событий — вычет по теории множеств, невозможному событию — пустое множество, обязательному событию — все пространство событий Ω.

σ-АЛГЕБРА (СИГМА-АЛГЕБРА)

В теории Колмогорова обычно не каждое подмножество пространств событий считается событием (более глубокие математические причины таких ограничений здесь приводиться не будут), но требуется, чтобы суммы и разности событий также являлись событиями и пустое множество и целое пространство событий Ω были событиями. Точнее, выделяется система S подмножеств пространства событий Ω, для которой установлены следующие условия:

1. Пустое множество и полное пространство событий Ω должны быть элементами системы S.

2. Если события A_1, A_2, ... являются элементами системы S, то $A_1 + A_2 \dots$ и $A_1 - A_2$ также должны быть элементами системы S. Система множеств, удовлетворяющая условиям 1, 2, называется σ-алгеброй. Следовательно, событиями являются такие подмножества Ω, которые являются элементами σ-алгебры системы S.

В некоторых случаях система S содержит все подмножества пространства событий Ω, а в других — лишь определенные подмножества Ω (например, только няблюдаемые события зачисляются в элементы системы S). Если эксперимент состоит в подбрасывании двух одинаковых кубиков, то из всех возможных 2^{36} подмножеств только 2^{21} подмножеств представляют собой наблюдаемые события, то есть те множества, которые содержат пару чисел (k, j) вместе со всеми парами чисел (j, k).

МАТЕМАТИЧЕСКОЕ ПОНЯТИЕ ВЕРОЯТНОСТИ

Аксиоматическим путем математическое понятие вероятности введено следующим образом. К каждому элементу σ-алгебры системы S, т. е. каждому событию, связанному с данным экспериментом, предписывается число $P(A)$, называнное вероятностью данного события, для которого устанавливаются следующие ограничения:

1. $$0 \leqq P(A) \leqq 1$$

(вероятность любого события находится в границах от 0 до 1).

2. $$P(\Phi) = 0, \qquad P(\Omega) = 1$$

(вероятность невозможного события 0, вероятность обязательного события 1).

3. Если $A_i \in S$ и $A_i A_j = \Phi$ $(i \neq j)$, то

$$P(A_1 + A_2 + \dots) = P(A_1) + P(A_2) + \dots,$$

то есть, вероятность суммы попарно исключающих друг друга событий совпадает с суммой вероятностей этих событий.

Данные 3 условия называются системой аксиом теории вероятности по Колмогорову.

Отметим, что наглядное представление вероятности (относительная частота при большом количестве экспериментов) удовлетворяет настоящим аксиомам. На основе этих аксиом уже может быть составлена математическая теория вероятности с применением чисто математических способов.

КОМБИНАТОРНЫЕ МЕТОДЫ ДЛЯ ОПРЕДЕЛЕНИЯ ВЕРОЯТНОСТЕЙ

Пусть проводится эксперимент, возможное количество результатов которого составляет n, и все n результатов одинаково вероятны.

Например, при подбрасывании игральной кости имеются 6 возможных результатов и все они действительно являются одинаково вероятными; при подбрасывании монеты имеются две возможности (орел или решка) они также одинаково вероятны. При подбрасывании двух игральных костей имеются 36 одинаково вероятных событий (при условии, что игральные кости или монеты, примененные в экспериментах, имеют одинаковые характеристики).

Очевидно, что если имеется n одинаково вероятных исходов некоторого эксперимента, то вероятность любого из исходов будет составлять $\frac{1}{n}$.

Если существует вероятность события A, которое является суммой k событий из событий $A_1, A_2, \ldots, A_n$, то есть имеет k различных возможных исходов, то получится, что

$$\boxed{P(A) = \frac{k}{n}.}$$

Если в некотором эксперименте число всех возможных событий составляет n, а k из них является благополучным с точки зрения исхода некоторого события A, то вероятность события A будет

$$\boxed{P(A) = \frac{k}{n} = \frac{\text{число благополучных случаев}}{\text{число всех случаев}},}$$

предполагая, что все случаи являются одинаково вероятными.

ПРИМЕРЫ

Пример 1. Пусть подбрасываются 2 одинаковых игральных кости. Какова вероятность того, что сумма чисел, которые выпали на двух игральных костях, больше чем 9? Количество всех возможных случаев — 36. Могут возникнуть следующие варианты:

$$
\begin{array}{cccccc}
(1,1) & (2,1) & (3,1) & (4,1) & (5,1) & (6,1)\\
(1,2) & (2,2) & (3,2) & (4,2) & (5,2) & (6,2)\\
(1,3) & (2,3) & (3,3) & (4,3) & (5,3) & (6,3)\\
(1,4) & (2,4) & (3,4) & (4,4) & (5,4) & (6,4)\\
(1,5) & (2,5) & (3,5) & (4,5) & (5,5) & (6,5)\\
(1,6) & (2,6) & (3,6) & (4,6) & (5,6) & (6,6).
\end{array}
$$

Из них благополучными являются варианты (4,6), (5,5), (5,6), (6,4), (6,5), (6,6). Следовательно, вероятность будет: $6/36 = 1/6$.

Можно предположить, что сумма на двух игральных костях может быть 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, следовательно имеется 11 возможных случаев, из которых 3 являются благополучными. Таким образом, значение вероятности составляет $\frac{3}{11}$. Такое рассуждение ошибочно, так как эти 11 возможных случаев не являются равновероятными. Например, вероятность выпадения 2-х очков существенно меньше, чем вероятности суммы выпадения 8 очков; 2 очка получаются только в случае выпадения 1 на каждой игральной кости; а 8 очков могут быть результатом одной из следующих комбинаций (2,6), (3,5), (4,4) или (6,2).

Пример 2. Какова вероятность того, что, играя в лото, можно угадать пять цифр? Посмотрим, сколько различных вариантов существует, в которых из 90 цифр могут быть выбраны пять (венгерская система лото). Видимо, $\binom{90}{5} = 43\,949\,268$ вариантов, из них только 1 является благополучным (когда правильно угаданы все 5 цифр). Таким образом, искомая вероятность составляет $\frac{1}{\binom{90}{5}} = 0{,}0000000228 = 2{,}28 \cdot 10^{-8}$. Возникает вопрос, являются ли все варианты одинаково вероятными между собой. Некоторые предполагают, что менее вероятно, что результатом розыгрыша лото была последовательность цифр 1, 2, 3, 4, 5, чем некоторая «нерегулярная» последовательность из 5 цифр, например 7, 8, 19, 40, 72. Обычно при доказательстве такого установления ссылаются на то, что в Венгрии в ходе розыгрышей лото каждый раз получалась такая «нерегулярная» последовательность из 5 цифр. Причина такого явления в том, что количество регулярных последовательностей намного меньше, чем количество нерегулярных последовательностей. Итак, если бы следовало угадать предварительно, регулярная или нерегулярная последовательность станет результатом розыгрыша, то естественно было бы более правильным предположить, что результатом розыгрыша будет нерегулярная последовательность. Учитывая, что цифры приходится угадывать точно, никакая последовательность из 5 цифр не имеет преимущества по сравнению с другими. По сути дела, предпочтение нужно было бы отдать регулярной последовательности цифр, чем нерегулярной (от этого вероятность выигрыша не изменится).

Но выигрыш в последовательности из 5 цифр, очевидно, будет больше, так как игроки обычно редко предполагают возможность рзыгрыша такой последовательности.

Подобная дискуссия будет определенной только в том случае, если определить, что подразумевается под термином «регулярная последовательность из 5 цифр».

Пример 3. Какова вероятность того, что при заполнении одного билета лото можно угадать 4 цифры из 5? Количество всех случаев здесь $\binom{90}{5}$.

В таком случае можно считать благополучным, если в розыгрыше будет угадано 4 цифры из 5, а пятая цифра будет одна из оставшихся 85 цифр (в Венгрии лото имеет всего 90 цифр от 1 до 90); поэтому количество благополучных случаев составляет $\binom{5}{4}\binom{85}{1}$, и искомая вероятность будет:

$$\frac{\binom{5}{4}\binom{85}{1}}{\binom{90}{5}} = \frac{425}{\binom{90}{5}} = \frac{425}{43\,949\,268} = 0{,}00000967 = 9{,}67 \cdot 10^{-6}.$$

Отметим, что сейчас нельзя было считать благополучным тот случай, когда угаданы все 5 цифр, поставленных нами. Если бы стоял вопрос, какова вероятность того, что можно угадать по крайней мере 4 цифры, то, поскольку количество благополучных случаев составляет $1 + \binom{5}{4}\binom{85}{1}$, искомая вероятность в данном случае будет:

$$\frac{\binom{5}{4}\binom{85}{1} + 1}{\binom{90}{5}} = \frac{426}{43\,949\,268} = 0{,}00000969 = 9{,}69 \cdot 10^{-6}.$$

Пример 4. Какова вероятность того, что при заполнении одного билета «тото» (вид спортивного пари в Венгрии) наугад удастся угадать все 13 результатов футбольных матчев. Результаты могут быть следующие: 1 — если выиграла первая команда; 2 — вторая, x — если матч закончился в ничью.

Так как имеется 3^{13} возможных вариантов заполнения билетов тото (вариации с повторениями), то количество всех случаев составляет 3^{13}. Очевидно, что один из них является благополучным.

Следовательно, вероятность выигрыша при заполнении билета наугад будет очень мала: $\frac{1}{3^{13}} = \frac{1}{2\,250\,423} = 0{,}000000444 = 4{,}44 \cdot 10^{-7}$.

Но при достаточной компетенции в данном вопросе можно угадать результаты всех встреч с существенно большей вероятностью.

УСЛОВНАЯ ВЕРОЯТНОСТЬ

Можно задать такой вопрос: какова вероятность наступления некоторого события A при условии, что уже произошло некоторое другое событие B. Такая вероятность обозначается символом $P(A|B)$.

Например: Если событием A является то, что завтра будет дождь, то какова вероятность, что дождь выпадет завтра, если утром будет облачная погода (событие B). Или, какова вероятность того, что в некотором объеме товара количество брака составляет меньше 3% (событие A), при условии, что среди первых 10-ти подвергшихся контролю штук товара не было брака (событие B).

Предположим, что проведено n экспериментов, в результате которых событие A произошло k_A раз, событие B — k_B раз и оба события AB — k_{AB} раз (см. указанные выше примеры). Тогда относительная частота события A составляет k_A/n в то время, как относительная частота события A среди тех экспериментов, когда и событие B произошло (условная относительная частота события при выполнении условия B):

$$\frac{k_{AB}}{k_B} = \frac{\frac{k_{AB}}{n}}{\frac{k_B}{n}} \sim \frac{P(AB)}{P(B)}.$$

Поэтому, по определению

$$P(A|B) = \frac{P(AB)}{P(B)}.$$

Данное соотношение имеет смысл только при условиии $P(B) > 0$.

Часто условная вероятность рассчитывается легко, и на основе знания условной вероятности (или условных вероятностей) следует определить вероятность самого события. Этому помогает следующая теорема:

ТЕОРЕМА О ПОЛНОЙ ВЕРОЯТНОСТИ

Пусть $B_1, B_2, \ldots, B_n$ — последовательность некоторых событий, в отношении которых имеет место утверждение, что из событий $B_1, B_2, \ldots, B_n$ любые два события взаимно исключают друг друга, а сумма событий B_i является обязательным событием. Выразим это в виде формулы:

$$B_1 + B_2 + \ldots + B_n = \Omega$$

и

$$B_iB_j = \Phi, \quad \text{если} \quad i \neq j.$$

Такая система событий $B_1, B_2, \ldots, B_n$ называется полной системой событий.

Пусть далее задано некоторое любое событие A. Тогда действительно следующее событие:

$$\boxed{P(A) = P(A|B_1)P(B_1) + P(A|B_2)P(B_2) + \ldots + P(A|B_n)P(B_n).}$$

Данное утверждение обычно называется *теоремой о полной вероятности*. Теорема может быть доказана следующим образом:

Поскольку $\Omega = B_1 + B_2 + \ldots + B_n$, то если перемножить это соотношение на A, получится, что

$$A = A\Omega = AB_1 + AB_2 + \ldots + AB_n.$$

Но события AB_i взаимно исключают друг друга, и поэтому действительно следующее соотношение:

$$P(A) = P(AB_1) + P(AB_2) + \ldots + P(AB_n).$$

По определению условной вероятности

$$P(AB)_i = P(A|B_i)P(B_i).$$

Таким образом, если подставить данное соотношение в приведенную выше формулу, получится теорема.

Возникает вопрос: если A и B представляют собой два любых события, то какая связь может быть установлена между условными вероятностями $P(A|B)$ и $P(B|A)$. Ответом может стать следующая формула:

$$P(B|A) = \frac{P(BA)}{P(A)} = \frac{P(AB)}{P(A)} = \frac{P(AB)}{P(B)} \cdot \frac{P(B)}{P(A)} = \frac{P(A|B)P(B)}{P(A)}.$$

Полезный результат будет достигнут, если в это соотношение подставить выражение

$$P(A|B_1)P(B_1) + P(A|B_2)\,P(B_2) + \ldots + P(A|B_n)P(B_n)$$

вместо $P(A)$ на основании теоремы о полной вероятности. Результат, полученный таким образом, обычно называется *теоремой Байеса*, и она доказывается следующим образом.

ТЕОРЕМА БАЙЕСА

Пусть $B_1, B_2, \ldots, B_n$ представляют собой полную систему событий, а A — любое событие. Тогда действительно следующее соотношение:

$$\boxed{P(B_i|A) = \frac{P(A|B_i)P(B_i)}{\sum\limits_{k=1}^{n} P(A|B_k)P(B_k)}.}$$

НЕЗАВИСИМОСТЬ

Два события A и B называются независимыми, если при наступлении одного из них вероятность наступления другого не изменяется. Следует подчеркнуть, что данное ограничение является более сильным, чем причинная независимость, т. е. из того, что одно событие не является причиной другого, не следует, что они независимы в вышеприведенном смысле.

Чтобы объяснить данное понятие, рассмотрим такой пример. Предположим, подбрасывается игральная кость; A представляет собой такое событие, что игральная кость выпадет на четное число, а B — такое событие, что она выпадет на число больше, чем 3. В этом случае между событиями A и B нет причинной зависимости, но в то время, как вероятность события A составляет 1/2, вероятность того, что произойдет событие A при условии осуществления события B (значит, игральная кость выпала на число, большее чем 3) — 2/3 ($\neq$ 1/2). Следовательно, события A и B не являются независимыми.

Для того, чтобы ввести точное понятие независимости, исходим из следующих рассуждений.

Пусть A и B представляют собой два независимых события; проводится n экспериментов; из них событие A произошло k_A раз, событие B—k_B раз, и оба события AB — k_{AB} раз. Как было отмечено, принцип независимости означает: наступление события A не делает более или менее вероятным наступление события B. Это значит, что относительная частота события A среди тех экспериментов, в которых произошло и событие B (т. е. среди k_B экспериментов), должна быть примерно такая же, как и во всей последовательности экспериментов $\left(\text{примерно столько, сколько значение } \dfrac{k_A}{n}\right)$.

Таким образом, $\dfrac{k_{AB}}{n} \sim \dfrac{k_A}{n} \cdot \dfrac{k_B}{n}$ (символ $\sim$ означает «примерно одинаково»). В соответствии с этим события A и B называются независимыми, если выполняется соотношение

$$P(AB) = P(A)P(B).$$

События $A_1, A_2, \ldots, A_k$ называются полностью независимыми, если для любых k событий ($k = 2,3, \ldots, n$) выполняется условие

$$P(A_{i_1}A_{i_2} \ldots A_{i_k}) = P(A_{i_1})P(A_{i_2}) \ldots P(A_{i_k})$$
$$(i_1 < i_2 < \ldots < i_k).$$

Если события A и B являются независимыми, то $P(A|B) = P(A)$.

СЛУЧАЙНАЯ ПЕРЕМЕННАЯ

В большинстве случаев на практике результат эксперимента может быть охарактеризован единственным числом. Иными словами, каждому возможному результату эксперимента (каждому элементарному событию) предписывается одно действительное число, т. е. в пространстве событий Ω задается функция $\xi(\omega)$ $(\omega \in \Omega)$ с действительными значениями. (Бывает, что различным элементарным событиям предписывается одно и то же число.)

В связи с определенными таким образом функциями обычно задается такое условие, что функция должна быть измеримой, т. е. множество элементарных событий (с обозначением ω), в котором функция $\xi(\omega)$ приобретает значение, меньшее, чем заданное число x, должно быть элементом σ-алгебры системы S. Суть этого условия заключается в том, что такое «событие», что значение функции $\xi(\omega)$ должно быть меньше, чем некоторое заданное число, действительно должно быть событием (должно быть элементом системы S), то есть должно иметь вероятность. Измеримая функция, определенная в пространстве событий ω, называется случайной переменной.

Иными словами, случайная переменная представляет собой численное значение, зависящее от случайности (например, уровень воды в Дунае в данный момент времени представляет собой случайную переменную).

ФУНКЦИЯ РАСПРЕДЕЛЕНИЯ И ФУНКЦИЯ ПЛОТНОСТИ

Свойства некоторой случайной переменной характеризуются ее функцией распределения. Функцией распределения задается вероятность того, что данная случайная переменная должна быть меньше некоторого числа x.

Точнее говоря, под определением функции распределения случайной переменной ξ подразумевается такая функция $F_\xi(x) = F(x)$ с действительным аргументом, с помощью которой задается вероятность того, что $\xi < x$, т. е.

$$F(x) = P(\xi < x)\,.$$

Легко убедиться в том, что функция распределения имеет следующие свойства:

1. $F(-\infty) = \lim\limits_{x\to-\infty} F(x) = 0,\ F(+\infty) = \lim\limits_{x\to+\infty} F(x) = 1,$
2. $F(x)$ растет монотонно, т. е. $F(x) \leq F(y)$, если $x < y$,
3. $F(x)$ является непрерывной слева, т. е. $F(x-0) = \lim\limits_{h\to 0} F(x-h) = F(x)\,.$

При знании функции распределения легко определяется вероятность и такого события, что случайная переменная ξ должна находиться в некотором заданном интервале, т. е.:

$$\boxed{\begin{aligned} &P(a \leqq \xi < b) = F(b) - F(a) \\ &P(a < \xi < b) = F(b) - F(a+0) \\ &P(a < \xi \leqq b) = F(b+0) - F(a+0) \\ &P(a \leqq \xi \leqq b) = F(b+0) - F(a) \,. \end{aligned}}$$

Легко увидеть, что вероятность такого события, что случайная переменная ξ равняется некоторому заданному числу a, совпадает со значением скачка функции распределения в некоторой точке a, то есть

$$\boxed{P(\xi = a) = F(a+0) - F(a) \,.}$$

Следовательно, если функция распределения $F(x)$ случайной переменной ξ представляет собой непрерывную функцию, то ξ приобретает любое возможное значение с вероятностью 0. Если функция $F(x)$ представляет собой скачкообразную функцию, то возможные значения ξ являются местами скачков в функции $F(x)$ (т. е. они представляют собой исчислимое множество). В таком случае ξ называется дискретной случайной переменной.

Если ξ представляет собой дискретную величину, а ее возможные значения — числа $x_1, x_2, \ldots$ и $P(\xi = x_i) = F(x_i + 0) - F(x_i) = p_i$, то последовательность чисел $p_1, p_2, \ldots$ называется распределением вероятностей величины ξ. (Естественно, что $p_1 + p_2 + \ldots = 1$.)

Среди функций распределения особенно важное значение имеют абсолютно непрерывные функции распределения, т. е. те, для которых можно найти функцию $f(x)$, в отношении которой выполняется условие

$$F(x) = \int_{-\infty}^{x} f(t)dt \,.$$

Такая функция $f(x)$ называется функцией плотности случайной переменной ξ.

Легко убедиться в том, что функция плотности $f(x)$ характеризуется следующими свойствами:

$$\boxed{\begin{aligned} &1.\ f(x) \geqq 0 \\ &2.\ \int_{-\infty}^{\infty} f(x)dx = 1. \end{aligned}}$$

При знании функции плотности легко проводить расчет для определения вероятности такого события, при котором случайная переменная ξ должна находиться в некотором заданном интервале. В частности,

$$P(a \leq \xi < b) = P(a < \xi < b) = P(a < \xi \leq b) = P(a \leq \xi \leq b) = \int_a^b f(t)dt .$$

В области случайных переменных понятие независимости может быть определено следующим образом. Случайные переменные $\xi_1, \xi_2 \ldots, \xi_n$ называются полностью независимыми, если для любой системы действительных чисел $x_1, x_2, \ldots, x_n$ действительно следующее соотношение:

$$P(\xi_1 < x_1, \xi_2 < x_2, \ldots, \xi_n < x_n) = P(\xi_1 < x_1)P(\xi_2 < x_2) \ldots P(\xi_n < x_n) .$$

Бесконечное количество случайных переменных $(\xi_1, \xi_2, \ldots)$ называется полностью независимым, если n любых из этих случайных переменных являются независимыми.

МАТЕМАТИЧЕСКОЕ ОЖИДАНИЕ

Два игрока играют в следующую игру: они подбрасывают игральную кость и второй игрок платит первому игроку столько рублей, какое число показывает игральная кость. Первый игрок перед каждым броском передает второму некоторую предварительно определенную зафиксированную сумму, например x рублей. Вопрос: каким образом следует определить значение x для того, чтобы игра была справедливой?

Необходимо определить также, что следует подразумевать под определением справедливая игра.

Игра называется справедливой, если при большом количестве партий средний выигрыш обоих игроков за одну партию будет примерно 0.

Определим, какой будет средний выигрыш первого игрока при такой игре. Предположим, что из n партий игральная кость выпадет k_1 раз на 1, k_2 раз на 2, k_3 раз на 3, k_4 раз на 4, k_5 раз на 5, k_6 раз на 6. Тогда первый игрок получит всего $1k_1 + 2k_2 + 3k_3 + 4k_4 + 5k_5 + 6k_6$ рублей, но, поскольку в каждой партии оплачивает x рублей, его суммарный выигрыш составит $k_1 + 2k_2 + 3k_3 + 4k_4 + 5k_5 + 6k_6 - nx$ рублей. Таким образом, средний выигрыш первого игрока за одну партию после n-ной партии будет

$$\frac{k_1 + 2k_2 + 3k_3 + 4k_4 + 5k_5 + 6k_6 + nx}{n} = \frac{k_1}{n} + 2\frac{k_2}{n} + 3\frac{k_3}{n} + 4\frac{k_4}{n} + 5\frac{k_5}{n} + 6\frac{k_6}{n} - x .$$

Число $\frac{k_1}{n}$ представляет собой относительную частоту выпадения единицы, которая ири достаточно большом n равна 1/6. Аналогично, значения

$$\frac{k_2}{n}, \quad \frac{k_3}{n}, \quad \frac{k_4}{n}, \quad \frac{k_5}{n}, \quad \frac{k_6}{n}$$

также будут $\approx 1/6$, поэтому

$$\frac{k_1}{n}+2\frac{k_2}{n}+3\frac{k_3}{n}+4\frac{k_4}{n}+5\frac{k_5}{n}+6\frac{k_6}{n}-x\sim$$

$$\sim\frac{1}{6}(1+2+3+4+5+6)-x=3{,}5-x\,.$$

Таким образом, игра будет справедлива, если $x=3{,}5$.

Предположим, что в некоторой игре с двумя игроками первый игрок может выиграть $x_1, x_2, \ldots, x_k$ рублей (среди чисел $x_1, x_2, \ldots, x_k$ может быть и отрицательное число), и вероятность того, что первый игрок выиграет x_i рублей, составляет p_i. Определим, каким будет средний выигрыш первого игрока на одну партию после достаточно большого количества партий.

Как и в предыдущем случае, средний выигрыш будет примерно

$$x_1p_1+x_2p_2+\ldots+x_kp_k\,.$$

Величина $x_1p_1+x_2p_2+\ldots+x_kp_k$ называется *математическим ожиданием* выигрыша первого игрока.

Игра называется справедливой, если

$$x_1p_1+x_2p_2+\ldots+x_kp_k=0\,;$$

игра называется выгодной для первого игрока, если

$$x_1p_1+x_2p_2+\ldots+x_kp_k>0\,,$$

а невыгодной, если

$$x_1p_1+x_2p_2+\ldots+x_kp_k<0\,.$$

Пусть ξ — любая дискретная случайная переменная, которая может быть равна любому из чисел $x_1, x_2, \ldots, x_k$, и вероятность того, что значение случайной величины ξ будет x_i, составляет именно p_i, т. е.

$$P(\xi=x_i)=p_i\,.$$

Обычно считают, что математическим ожиданием случайной переменной ξ является выражение

$$M(\xi)=\sum_{i=1}^{k}p_ix_i$$

(обозначается через $M(\xi)$, в литературе на английском языке — через $E(\xi)$).

Понятие математического ожидания случайной переменной ξ можно определить не только в том случае, если ξ представляет собой дискретную величину.

В общем случае математическое ожидание задается формулой

$$M(\xi) = \int_{-\infty}^{+\infty} x dF(x),$$

(если предположить, что данный интеграл существует. Если такой интеграл не существует, то считается, что не существует математическое ожидание случайной переменной ξ.)

Легко заметить, что в дискретном случае данное соотношение приводится к уже знакомому нам соотношению. Можно показать, что такое обобщение является естественным обобщением понятия математического ожидания, введенного на дискретный случай.

В случае, если $F(x)$ представляет собой абсолютно непрерывную функцию, математическое ожидание определяется следующим соотношением:

$$\boxed{M(\xi) \int_{-\infty}^{+\infty} x dF(x) = \int_{-\infty}^{+\infty} x f(x)\, dx.}$$

Из свойств математического ожидания здесь упомянем следующие:

$$\boxed{M(a\xi + b\eta + c) = aM(\xi) + bM(\eta) + c,}$$

где a, b, c — вещественные числа. Если ξ и η являются независимыми друг от друга, то справедливо следующее соотношение:

$$\boxed{M(\xi\eta) = M(\xi)\, M(\eta).}$$

ДИСПЕРСИЯ

Пусть результатом эксперимента и в этом случае будет некоторое число ξ. Математическим ожиданием ξ определяется, каково будет среднее значение полученных чисел, если много раз повторять данный эксперимент. Результаты самого эксперимента колеблются вокруг данного среднего значения. Мерой этих колебаний является *дисперсия*. Определение дисперсии задается в следующем виде:

Рассмотрим отклонение результата эксперимента, значения ξ от значения математического ожидания, т. е. от величины $M(\xi)$, рассмотрим число

$\xi - M(\xi)$. Оно может быть или отрицательным, или положительным. Так как для определения величины колебаний необходимо знать не знак, а величину отклонения, — рассматривается величина $(\xi - M(\xi))^2$. На основании такой величины дисперсией называется квадратный корень значения математического ожидания величины $(\xi - M(\xi))^2$. Дисперсия обычно обозначается символом $D(\xi)$ и определяется следующим соотношением:

$$D(\xi) = \sqrt{M[(\xi - M(\xi))^2]}.$$

Расчет дисперсии можно проводить проще, с помощью следующих несложных соотношений:

$$\boxed{\begin{gathered} M[(\xi - M(\xi))^2] = M(\xi^2) - M^2(\xi), \\ M(\xi^2) = \int_{-\infty}^{+\infty} x^2 dF(x). \end{gathered}}$$

Пусть $\xi_1, \xi_2, \ldots, \xi_n$ случайные переменные с одинаковым распределением и с дисперсией, равной σ. Среднеквадратичное отклонение случайной суммы $\xi_1 + \xi_2 \ldots + \xi_n$ составляет $n\sigma^2$, а дисперсия — $\sqrt{n}\sigma$.

Если ξ представляет собой дискретную случайную величину с возможными значениями $x_1, x_2, \ldots$ и с распределением $p_1, p_2, \ldots$, то действительно следующее соотношение:

$$\boxed{M(\xi^2) = \sum_{k=1}^{\infty} x_k^2 p_k.}$$

В том случае, когда функция распределения случайной величины ξ представляется как абсолютно непрерывная функция с функцией плотности $f(x)$, то имеет место следующее соотношение:

$$\boxed{M(\xi^2) = \int_{-\infty}^{+\infty} x^2 f(x)\, dx.}$$

Кроме понятий математического ожидания и дисперсии исследуются и понятия моментов высшего порядка.

Под определением n-ого момента случайной переменной ξ подразумевается математическое ожидание $M(\xi^n)$ случайной переменной ξ^n. Аналогичным образом под определением n-ого центрального момента случайной переменной ξ подразумевается значение $M((\xi - M(\xi))^n)$ n-ого абсолютного момента $-M(|\xi|^n)$, n-ого абсолютного центрального момента $-M(|\xi - M(\xi)|^n)$.

КОРРЕЛЯЦИЯ

Как упоминалось выше, если ξ и η представляют собой две независимые случайные переменные, то $M(\xi\eta) = M(\xi)\,M(\eta)$, что одновременно означает: $M\big((\xi - M(\xi))\,(\eta - M(\eta))\big) = 0$. В случае зависимых переменных из величины $M\big((\xi - M(\xi))\,(\eta - M(\eta))\big)$ можно было бы сделать вывод об интенсивности связи между переменными ξ и η. Но бывают такие случаи, когда $M\big((\xi - M(\xi))\,(\eta - M(\eta))\big) = 0$, но переменные ξ и η не являются независимыми.

Пусть ξ и η — две зависимые случайные переменные. Предположим, что связь между ними носит такой характер, что если ξ принимает достаточно большое значение (большее, чем математическое ожидание самой этой величины), то, вероятно, и значение η будет большое. (Например, если ξ означает высоту случайно избранного лица, а η — его вес, то обычно большим значениям величины η соответствуют большие значения величины η.) Это значит, что в случае такой зависимости произведение $(\xi - M(\xi))\,(\eta - M(\eta))$ обычно является положительным, но $M\big((\xi - M(\xi))\,(\eta - M(\eta))\big) > 0$. Если же связь между переменными ξ и η является обратной по сравнению с предыдущей (большим значениям ξ соответствуют маленькие значения η), действительно следующее соотношение:

$$M\big((\xi - M(\xi))\,(\eta - M(\eta))\big) < 0.$$

Величина $M\big((\xi - M(\xi))\,(\eta - M(\eta))\big)$ зависит не только от интенсивности связи между переменными ξ и η, но и от величины дисперсии случайных переменных ξ и η. Таким образом,

для измерения интенсивности связи между случайными переменными ξ и η введено понятие корреляции:

$$R(\xi,\,\eta) = \frac{M\big((\xi - M(\xi))\,(\eta - M(\eta))\big)}{D(\xi)\,D(\eta)} = \frac{M(\xi\eta) - M(\xi)\,M(\eta)}{D(\xi)\,D(\eta)}.$$

Легко проверить следующие два свойства корреляции $R(\xi,\,\eta)$:

1. $-1 \leqq R(\xi,\,\eta) \leqq +1.$
2. $R(\xi,\,\eta) = 0,$ если ξ и η являются независимыми.

На основе рассуждений, представленных выше, что корреляция $R(\xi,\,\eta)$ принимает большое положительное значение (около $+1$), можно сделать вывод, что если величина ξ «велика» (больше, чем ее математическое ожидание), то величина η также велика. Таким образом, величины ξ и η *сильно положительно коррелированы* между собой. Аналогичное утверждение может быть и в том случае, если значение корреляции $R(\xi,\,\eta)$ близко к -1.

Если $R(\xi,\,\eta)$ равно 0, то считается, что величины ξ и η являются *некоррелированными*. Обычно это означает, что величины ξ и η мало зависят друг от друга, но они не обязательно независимы.

РАСПРЕДЕЛЕНИЯ ВЕРОЯТНОСТИ

Ниже рассматриваются несколько конкретных функций распределения, часто встречающихся на практике.

БИНОМИАЛЬНОЕ РАСПРЕДЕЛЕНИЕ

Предположим, что из 10 000 предметов потребления (например, электрических ламп) случайным образом (например, путем розыгрыша), выбираются 100 штук для испытания (проба из 100 элементов). Какова вероятность того, что количество бракованных предметов в пробе составляет k (k может быть $0, 1, 2, \ldots, 100$), если среди 10 000 предметов количество бракованных предметов — 300?

Отметим, что на практике возникает обратный вопрос, то есть если известно количество бракованных предметов в пробе, то что можно сказать о количестве бракованных предметов во всей партии товара?

Данная задача решается таким образом, что из 10 000 предметов выбирается один предмет путем случайного отбора; рассматривается, является ли данный предмет бракованным, а затем предмет кладется обратно. После этого выбирается следующий элемент пробы. Может случиться, что один и тот же элемент испытывается дважды. Если количество элементов в пробе мало по сравнению с общим количеством всех предметов, то не имеет принципиального значения, выбирается ли проба с возвратом или без возврата предмета, так же, как мала вероятность того, что один элемент испытывается дважды.

Вернемся к решению задачи. Вероятность того, что электрическая лампа, выбранная случайно, является бракованной, составляет $\frac{300}{10\,000} = \frac{3}{100} = 0{,}03$; вероятность того, что она не бракована — 0,97. Таким образом, легко ответить на вопрос: какова вероятность того, что в пробе не будет ни одной бракованной лампы или что каждый элемент пробы будет бракованный. Вероятность того, что ни одна лампа не будет бракованная, — $(0{,}97)^{100}$, так как вероятность того, что первый элемент пробы не будет бракованный — 0,97. Таким же образом, вероятность того, что второй элемент из пробы не будет бракованный — 0,97 (поскольку первый элемент кладется обратно), и так как два выбора являются независимыми между собой, то вероятность того, что ни один из первых двух элементов пробы не будет бракованный, составляет $(0{,}97)^2$. Таким же образом, продолжая выбор элементов пробы, получится значение $(0{,}97)^{100}$. Аналогичным образом получится, что $(0{,}03)^{100}$ будет вероятность того, что каждый элемент пробы будет бракованный. Вероятность того, что *один* элемент будет бракованный, получится следующим образом. Имеется 100 различных вариантов возможности такого варианта: может ли первый выбранный элемент быть бракованным при условии, что все остальные 99 выбранных элементов будут небракованные; или — второй выбранный элемент будет бракованным и т. д.

Вероятность того, что первый выбранный элемент будет бракованный и все остальные 99 элементов пробы будут хорошими, составляет $0{,}03 \cdot (0{,}97)^{99}$ (из-за независимости). Так как вероятность всех вышеуказанных 100 событий одна и та же, искомая вероятность будет $100 \cdot 0{,}03 \cdot (0{,}97)^{99}$.

Аналогичным образом можно получить вероятность того, что в каждой пробе имеются два бракованных предмета, только следует учитывать, что имеются $\binom{100}{2}$ вариантов наступления такого события. Например, первый и второй, или первый и третий, или второй и пятый выбранные предметы будут бракованными. Вероятность того, что первый и второй элементы пробы будут бракованными, а остальные нет, будет:

$$(0{,}03)^2 \cdot (0{,}97)^{98}$$

и поэтому искомая вероятность:

$$\binom{100}{2} (0{,}03)^2 (0{,}97)^{98}.$$

Таким же образом можно показать, что вероятность того, что в пробе будет k бракованных предметов, составляет

$$\binom{100}{k} (0{,}03)^k (0{,}97)^{100-k}.$$

В общем случае, если в N предметах товара имеются M штук бракованных предметов, вероятность того, что в пробе из этого товара из n элементов находится k бракованных предметов, составляет:

$$\boxed{P_k = \binom{n}{k} \left(\frac{M}{N}\right)^k \left(1 - \frac{M}{N}\right)^{n-k}.}$$

ФОРМУЛА БЕРНУЛЛИ

Эту задачу можно решать в несколько обобщенном виде. Пусть A представляет собой событие с вероятностью p и пусть проводится n независимых экспериментов, результатом каждого из которых является наступление или ненаступление события A. Тогда вероятность того, что из n экспериментов событие A произойдет k раз, будет (формула Бернулли):

$$\boxed{P_k = \binom{n}{k} p^k (1 - p)^{n-k}.}$$

Числа $P_0, P_1, P_2, \ldots, P_n$ называются биномиальным распределением.

Математическое ожидание и дисперсия случайной переменной ξ биномиального распределения имеет следующий вид:

$$M(\xi) = \sum_{k=0}^{n} k \binom{n}{k} p^k (1-p)^{n-k} = \sum_{k=1}^{n} n \binom{n-1}{k-1} p^k (1-p)^{n-k} =$$

$$= np \sum_{k=1}^{n} \binom{n-1}{k-1} p^{k-1} (1-p)^{n-k} = np$$

и

$$D^2(\xi) = M(\xi^2) - M^2(\xi) = \sum_{k=0}^{n} k^2 \binom{n}{k} p^k (1-p)^{n-k} - (np)^2 = np(1-p).$$

В области практического применения биномиального распределения существенным препятствием является то обстоятельство, что расчет величины $\binom{n}{k}$ в формуле является громоздким, когда значения n и k велики.

Формула Муавра—Лапласа. По приведенной выше причине здесь приводится приближенная формула, с помощью которой трудоемкие расчеты, связанные с рядом задач в области биномиальных распределений, могут быть проведены в упрощенном виде. Данная формула известна в специальной литературе как формула Муавра—Лапласа.

Поскольку данная формула представляет собой частный случай теорем о центральных граничных распределениях, мы еще вернемся к данной проблеме.

Утверждение формулы Муавра—Лапласа имеет следующий вид:

$$\sum_{k=l}^{m} \binom{n}{k} p^k (1-p)^{n-k} \sim \Phi\left(\frac{m-np}{\sqrt{np(1-p)}}\right) - \Phi\left(\frac{l-np}{\sqrt{np(1-p)}}\right).$$

Здесь символ $\sim$ означает, что левая и правая части асимптотически совпадают при $n \to \infty$, при условии, что разница значений m и l с одной стороны и np с другой — не будет больше чем $O(\sqrt{n})$. Функция $\Phi(x)$ в формуле определяется с помощью следующего выражения:

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}} dt.$$

Интеграл в данной формуле нельзя вычислить элементарным путем. Ниже предложена сокращенная таблица его значений; в большинстве книг по теории вероятностей представлены более подробные таблицы значений функции $\Phi(x)$, $\binom{n}{k} p^k(1-p)^{n-k}$, $\sum_{k=r}^{n} \binom{n}{k} p^k(1-p)^{n-k}$.

x	$\Phi(x)$	x	$\Phi(x)$
0,00	0,5000	0,40	0,6554
0,05	0,5199	0,45	0,6736
0,10	0,5398	0,50	0,6915
0,15	0,5596	0,55	0,7088
0,20	0,5793	0,60	0,7257
0,25	0,5987	0,65	0,7422
0,30	0,6179	0,70	0,7580
0,35	0,6368	0,75	0,7734
0,80	0,7881	1,70	0,9554
0,85	0,8023	1,75	0,9599
0,90	0,8159	1,80	0,9641
0,95	0,8289	1,85	0,9678
1,00	0,8413	1,90	0,9713
1,05	0,8531	1,95	0,9744
1,10	0,8643	2,00	0,9772
1,15	0,8749	2,10	0,9821
1,20	0,8849	2,20	0,9861
1,25	0,8944	2,30	0,9893
1,30	0,9032	2,40	0,9918
1,35	0,9115	2,50	0,9938
1,40	0,9192	2,60	0,9953
1,45	0,9265	2,70	0,9965
1,50	0,9332	2,80	0,9974
1,55	0,9394	2,90	0,9981
1,60	0,9452	3,00	0,9986
1,65	0,9505	3,80	0,9999

Для применения таблицы необходимо знать, что функция $\Phi(x)$ является монотонно возрастающей и $0 \leq \Phi(x) \leq 1$, поэтому, если $x > 3{,}8$, то значение функции $\Phi(x)$ равно 1 с точностью до четырех знаков; при отрицательных значениях x значение функции $\Phi(x)$ получится из таблицы с помощью формулы:

$$\Phi(-x) = 1 - \Phi(x).$$

Примеры

применения формулы Муавра—Лапласа

Пример 1. Сколько раз следует подбрасывать монету для того, чтобы относительная частота количества выпадений решки была в диапазоне от 0,49 до 0,51 с вероятностью 0,95?

Если монета подбрасывается n раз, то нахождение относительной частоты количества выпадений решки в диапазоне от 0,49 до 0,51 означает, что количество решек колеблется между значениями $0{,}49n$ и $0{,}51n$.

Таким образом, вероятность того, что после n экспериментов относительная частота выпадений решки находится в диапазоне от 0,49 до 0,51, составляет:

$$\sum_{k=0,49n}^{k=0,51n} \binom{n}{k} \left(\frac{1}{2}\right)^n.$$

Следовательно, наша задача состоит в том, чтобы выбрать целое число n таким образом, чтобы удовлетворялось неравенство

$$\sum_{k=0,49n}^{0,51n} \binom{n}{k} \left(\frac{1}{2}\right)^n \geqq 0,95.$$

Для определения числа n может быть применена приближенная формула. Так как

$$\sum_{k=0,49n}^{0,51n} \binom{n}{k} \left(\frac{1}{2}\right)^n \sim \Phi\left(\frac{0,51n - 0,5n}{\sqrt{n(0,5)^2}}\right) - \Phi\left(\frac{0,49n - 0,5n}{\sqrt{n(0,5)^2}}\right) =$$

$$= \Phi\left(\frac{\sqrt{n}}{50}\right) - \Phi\left(-\frac{\sqrt{n}}{50}\right) = 2\Phi\left(\frac{\sqrt{n}}{50}\right) - 1,$$

поэтому было бы достаточно выбрать число n такое, чтобы удовлетворялось неравенство

$$2\Phi\left(\frac{\sqrt{n}}{50}\right) - 1 \geqq 0,95,$$

т. е.

$$\Phi\left(\frac{\sqrt{n}}{50}\right) \geqq \frac{1 + 0,95}{2} = 0,975.$$

Из таблицы видно, что (на основании интерполяции) $R(1,96) = 0,975$. Таким образом, неравенство, представленное выше, удовлетворяется, если

$$\frac{\sqrt{n}}{50} \geqq 1,96,$$

то есть, если

$$n \geqq (50 \cdot 1,96)^2 = (98)^2 = 9604.$$

Итак, если проводить 9604 (или более) экспериментов, то относительная частота решек будет в диапазоне от 0,49 до 0,51 с вероятностью по меньшей мере 95%; то есть вероятность того, что относительная частота отличается от вероятности $p = \frac{1}{2}$ больше чем на 0,01 при 9604 подбрасываниях, будет меньше, чем 5%.

Проводя любое количество экспериментов, нельзя быть *уверенным* в том, что относительная частота будет близка к вероятности. Результат, по которому при проведении 9604 экспериментов меньше, чем 5% вероятность того, что относительная частота отличается от вероятности на больше чем 0,01 может быть сформулирована следующим образом: если проводить много раз подряд такой эксперимент, что подбрасывается монета 9604 раза в каждом эксперименте, тогда в общем случае меньше, чем 5% от этих экспериментов представляет собой такую последовательность подбрасываний, в которой относительная частота отличается от значения 1/2 на больше чем 0,01. Если подбрасывать монету 9 604 000 раз и после каждого 9604-го подбрасывания проводить расчет на определение относительной частоты предыдущих 9604 подбрасываний, то можно ожидать, что из полученных таким образом 1000 значений относительной частоты 950 будут находиться в диапазоне от 0,49 до 0,51 и 50 значений будут находиться вне этих границ.

Пример 2. Часто бывает, что машины некоторых типов не работают постоянно, а включаются только в некоторые моменты. Так, например, если на каком-то заводе применяются 20 сварочных пистолетов, то всегда будет несколько из них, которые не включены в данный момент времени, так как, скажем, при резании отрезанный кусок металла следует отодвинуть и на это время пистолет приходится отложить. Таким же образом, если на каком-то заводе 100 рабочих занимаются точечной сваркой, то в промежуток между двумя сварками рабочие не пользуются своими инструментами.

Итак, возникает вопрос, сколько примерно инструментов используют в данный момент времени? Данный вопрос особенно важен. Ведь бетонорежущий пистолет работает с помощью сжатого воздуха, подаваемого компрессором. Использование компрессора такой мощности, с которой подается достаточное количество сжатого воздуха для 20 бетонорежущих пистолетов, является весьма дорогостоящим. Поэтому необходимо знать, какой мощности компрессор следует применять для того, чтобы для каждого *работающего* бетонорежущего пистолета хватало бы сжатого воздуха.

Предположим, например, что машина определенного типа всегда работает одну минуту без остановки и ею пользуются 20 раз в час. На заводе имеется 100 таких машин. Для какого числа машин следует обеспечивать энергию, для того, чтобы не было остановки с вероятностью 95%? Иначе говоря, каково наименьшее число n_0, при котором одновременно работает меньше, чем n_0 машин с вероятностью 95%.

Так как одна машина работает в среднем 20 минут в час, вероятность того, что в данный момент времени работает некоторая определенная машина, составляет $20/60 = 1/3$. Поэтому вероятность того, что в некоторый данный момент времени работает именно k машин из 100 (предполагая независимость):

$$\binom{100}{k}\left(\frac{1}{3}\right)^k\left(\frac{2}{3}\right)^{100-k}.$$

Вероятность того, что из 100 машин работают по крайней мере n_0:

$$\sum_{k=0}^{n_0} \binom{100}{k} \left(\frac{1}{3}\right)^k \left(\frac{2}{3}\right)^{100-k}.$$

Для решения нашей задачи достаточно найти такое наименьшее число n_0, для которого удовлетворяется неравенство

$$\sum_{k=0}^{n_0} \binom{100}{k} \left(\frac{1}{3}\right)^k \left(\frac{2}{3}\right)^{100-k} \geqq 0{,}95.$$

При определении числа n_0 можно опять воспользоваться приближенной формулой:

$$\sum_{k=0}^{n_0} \binom{100}{k} \left(\frac{1}{3}\right)^k \left(\frac{2}{3}\right)^{100-k} \sim \Phi\left(\frac{n_0 - 100\frac{1}{3}}{\sqrt{100\frac{1}{3}\frac{2}{3}}}\right) - \Phi\left(\frac{-100\frac{1}{3}}{\sqrt{100\frac{1}{3}\frac{2}{3}}}\right) =$$

$$= \Phi\left(\frac{n_0 - 33{,}3}{\frac{10}{3}\sqrt{2}}\right) - \Phi\left(\frac{-10}{\sqrt{2}}\right) = \Phi\left(\frac{n_0 - 33{,}3}{\frac{10}{3}\sqrt{2}}\right) - \Phi(-7{,}1) \sim \Phi\left(\frac{n_0 - 33{,}3}{3{,}33\sqrt{2}}\right).$$

Следовательно, число n_0 следует выбирать таким образом, чтобы удовлетворялось неравенство:

$$\Phi\,\frac{n_0 - 33{,}3}{3{,}33\sqrt{2}} \geqq 0{,}95.$$

Из таблицы находим, что

$$\Phi(1{,}64) = 0{,}9495$$

и

$$\Phi(1{,}65) = 0{,}9505,$$

поэтому, путем применения интерполяции,

$$\Phi(1{,}645) = 0{,}95.$$

Итак, должно удовлетворяться неравенство

$$\frac{n_0 - 33{,}3}{4{,}71} \geqq 1{,}645,$$

то есть,

$$n_0 \geqq 1{,}645 \cdot 4{,}71 + 33{,}3 = 41{,}1.$$

Результат показывает, что если из 100 машин всего для 42 обеспечивается достаточная энергия, то и в этом случае уже не будет остановки в производстве с вероятностью 95%, поскольку меньше, чем 42 машины работают одновременно с вероятностью 95%.

ГИПЕРГЕОМЕТРИЧЕСКОЕ РАСПРЕДЕЛЕНИЕ

Предположим, что из N предметов выбирается проба из n элементов без возврата. Какова вероятность того, что в пробе будет именно k бракованных предметов, предполагая, что во всем количестве предметов K было бракованных?

Из N предметов проба из n элементов может быть выбрана по $\binom{N}{n}$ различным вариантам. Предположим, что выбор любого n является одинаково вероятным. Необходимо знать, каково количество выгодных случаев, т. е. сколько вариантов имеется при выборе пробы из n элементов таким образом, чтобы в пробе было именно k бракованных предметов. Поскольку k штук из K бракованных предметов можно выбирать по $\binom{K}{k}$ различным вариантам, а $n-k$ штук хороших из $N-K$ хороших предметов — по $\binom{N-k}{n-k}$, то количество выгодных случаев составляет $\binom{K}{k}\binom{N-K}{n-k}$, и, таким образом, искомая вероятность будет

$$V_k = \frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}}.$$

Известно, что если величина N существенно больше величины n, то данная вероятность не отличается существенно от результата, полученного на основе проб с возвратом предметов, то есть:

$$\boxed{V_k = \frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}} \sim \binom{n}{k}\left(\frac{K}{N}\right)^k\left(1-\frac{K}{N}\right)^{n-k}.}$$

Прогрессия чисел $V_0, V_1, V_2, \ldots, V_n$ называется гипергеометрическим распределением.

Заметим, что математическое ожидание и дисперсия некоторой случайной переменной ξ с гипергеометрическим распределением рассчитываются с помощью следующих формул:

$$M(\xi) = n\frac{K}{N}$$

$$D(\xi) = \sqrt{\frac{Kn}{N}\left(1 - \frac{K}{N}\right)\left(1 - \frac{n-1}{N-1}\right)}.$$

ПУАССОНОВСКОЕ РАСПРЕДЕЛЕНИЕ

В некоторый магазин приходят покупатели в случайные моменты времени, начиная с момента открытия (с момента $t = 0$) магазина. Пусть обозначается через ξ_t количество покупателей, пришедших в магазин до момента времени t. Нас интересует распределение случайной переменной ξ_t, т. е. какова вероятность того, что до некоторого момента времени t именно k ($k = 0, 1, 2, \ldots$) покупателей придет в магазин. Следует определить функции $V_k(t) = P(\xi_t = k)$. Естественно, для того, чтобы определить функции $V_k(t)$, необходимо найти соответствующие условия.

Условиями являются следующие:

1. *Независимость*. Пусть $[a, b]$ и $[c, d]$ представляют собой два интервала времени, следующих друг за другом (т. е. $a < b \le c < d$). Предположим, что число покупателей $\xi_b - \xi_a$, пришедших в промежуток времени от a до b, независимо от числа покупателей $\xi_d - \xi_c$, пришедших в промежуток времени от c до d. В виде формулы это выражается следующим образом:

$$P(\xi_b - \xi_a = k,\ \xi_d - \xi_c = l) = P(\xi_b - \xi_a = k)\, P(\xi_d - \xi_c = l).$$

Данное условие действительно не представляет собой усиленное ограничение.

2. *Однородность*. Пусть $(a, a + d)$ и $(b, b + d)$ представляют собой два одинаковых интервала времени, следующих друг за другом. Предполагается, что вероятность того, что в промежуток времени от a до $a + d$ приходит k покупателей, совпадает с вероятностью того события, что k покупателей приходит в промежуток времени от b до $b + d$, т. е.

$$P(\xi_{b+d} - \xi_b = k) = P(\xi_{a+d} - \xi_a = k) = V_k(d).$$

Практически данное условие означает, что в магазине нет часов пик; в двух интервалах времени одинаковой длины приходит «примерно» одинаковое количество покупателей. Конечно, такое условие обычно не выполняется, но если рассматривается некоторый более короткий интервал времени (например, промежуток времени от 9 до 11 часов), то данное условие вообще уже выполняется.

3. *Условие ординарности.* Из данного условия вытекает, что если два покупателя приходят в магазин вместе, то они рассматриваются как один покупатель. Точнее, ставится такое условие, что вероятность того, что за очень короткий интервал времени приходит больше, чем один покупатель, по существу совпадает с вероятностью того, что приходит один или больше покупателей. Если выразить формулой:

$$\lim_{\Delta t \to 0} \frac{\sum\limits_{k=1}^{\infty} V_k(\Delta t)}{V_1(\Delta t)} = 1 .$$

Данное условие также не представляет собой слишком сильное ограничение.

На основе приведенных выше условий уже может быть определена форма функций $V_k(t)$ при рассмотрении следующей ситуации: k покупателей приходит до момента времени $t + \Delta t$, k покупателей пришли уже до момента времени t и ни один покупатель не приходит в промежуток времени от t до $t + \Delta t$; события могут происходить и таким образом, что $k - 1$ покупателей пришли до момента времени t и 1 покупатель приходит в промежуток времени от t до $t + \Delta t$ и т. д. Если выразить это в виде формулы, то получится следующее:

$$V_k(t + \Delta t) = \sum_{j=0}^{k} V_j(t)\, {}_{k-j}(\Delta t) .$$

Из данного соотношения путем применения простых преобразований получится следующая система дифференциальных уравнений:

$$\begin{aligned} V_0'(t) &= -\lambda V_0(t) \\ V_k'(t) &= -\lambda V_k(t) + \lambda V_{k-1}(t) , \qquad (k = 1, 2, \ldots) , \end{aligned}$$

где $\lambda = V_0'(0)$. Решение системы дифференциальных уравнений имеет следующий вид:

$$\boxed{P(\xi_t = k) = V_k(t) = \frac{(\lambda t)^k}{k!} e^{-\lambda t} ,}$$

откуда непосредственно могут быть получены математическое ожидание и дисперсия случайной переменной ξ_t:

$$\boxed{M(\xi_t) = \lambda t , \qquad D(\xi_t) = \sqrt{\lambda t} .}$$

Таким образом, оказывается, что среднее число покупателей, приходящих в единицу времени, составляет λ (математическое ожидание).

В конечном итоге можно сделать вывод, что покупатели приходят в магазин случайно, но при выполнении вышеприведенных трех условий среднее число покупателей, приходящих в магазин в единицу времени, составляет λ, и вероятность того, что k покупателей приходит в промежуток времени длины t составляет

$$V_k(t) = \frac{(\lambda t)^k}{k!} e^{-\lambda t}.$$

Прогрессия чисел $V_k(t)$ ($k = 0, 1, 2, \ldots$) называется пуассоновским распределением.

В природе часто встречается пуассоновское распределение (например, радиоактивный распад). Многие события, происходящие случайно во времени, обычно описываются с помощью пуассоновского распределения. Аналогичным образом могут быть приведены понятия пуассоновского распределения в плоскости или в пространстве, когда в плоскости или в пространстве точки распределены случайным образом. Звезды расположены в пространстве приблизительно по пуассоновскому распределению.

Интересно рассмотреть такой вопрос: насколько могут быть ослаблены три приведенные выше условия. Самым существенным является первое из трех условий, а если пренебречь вторым и третьим условиями, могут быть получены распределения, близкие к пуассоновскому и описиваемые математически без особого труда.

ПОКАЗАТЕЛЬНОЕ РАСПРЕДЕЛЕНИЕ

Первое рассматриваемое здесь непрерывное распределение представляет собой показательное распределение. Исходная задача ставится следующим образом: пусть ξ представляет собой интервал времени случайной длины. Например, ξ может означать длину телефонного разговора или срок жизни радиоактивного атома (длину интервала времени от его возникновения до распада), или время между прибытиями на остановку двух, следующих друг за другом автобусов. Задача заключается в определении распределения функции случайной величины ξ, то есть вероятности того, что данный промежуток времени меньше чем ξ.

К данной задаче ставится следующее условие. Если уже известно, что величина ξ больше, чем некоторое заданное число t_0, то вероятность того, что она одновременно больше, чем число $t_0 + \tau$, совпадает с вероятностью того, что величина ξ больше, чем τ (безусловно). Если составить выражение из данного условия, оно примет следующий вид:

$$P(\xi > t_0 + \tau | \xi > t_0) = P(\xi > \tau),$$

или с применением функции распределения:

$$F(t_0 + \tau) - F(t_0) = F(\tau)\bigl(1 - F(t_0)\bigr); \qquad (t \geqq 0;\ \tau \geqq 0). \tag{1}$$

Условие станет более очевидным, если заметить, что данное условие *не* выполнится, если ξ представляет собой срок жизни человека, так как в таком случае выполнение условия означало бы, что вероятность того, что младенец доживет до десятилетнего возраста ($\tau > 10$), равна вероятности того, что девяностолетний человек ($t_0 = 90$) доживет до возраста 100 лет (а ведь очевидно, что вероятность первого события существенно больше). Наше условие выполняется на срок жизни радиоактивных атомов и в достаточной степени действительно на длину телефонных разговоров, так как вероятность того, что человек, разговаривая уже 10 минут по телефону, закончит разговор в течение следующей минуты, примерно равняется вероятности того, что кто-то заканчивает телефонный разговор в первую минуту переговора.

Можно убедиться в том, что единственная функция распределения, для которой выполняется условие (1), имеет следующий вид:

$$F(t) = \begin{cases} 1 - e^{-\lambda t}, & \text{если} \quad t \geqq 0 \\ 0, & \text{если} \quad t < 0 \end{cases}$$

($\lambda > 0$). Такая функция распределения называется показательной функцией распределения.

Легко увидеть, что если ξ представляет собой случайную переменную с показательным распределением, то функцией плотности величины ξ является функция

$$\boxed{f(t) = \begin{cases} \lambda e^{-\lambda t}, & \text{если} \quad t \geqq 0 \\ 0, & \text{если} \quad t < 0; \end{cases}}$$

математическое ожидание величины ξ выражается так:

$$\boxed{M(\xi) = \int_0^\infty t\, f(t)\, dt = \frac{1}{\lambda};}$$

а ее дисперсия:

$$\boxed{D(\xi) = \sqrt{M(\xi^2) - M^2(\xi)} = \sqrt{\int_0^\infty t^2 \lambda e^{-\lambda t}\, dt - \frac{1}{\lambda^2}} = \frac{1}{\lambda}.}$$

НОРМАЛЬНОЕ РАСПРЕДЕЛЕНИЕ

Нормальное распределение является наиболее часто встречающимся в математической статистике непрерывным распределением. Его важность станет очевидной в связи с обсуждением теорем о центральных граничных распределениях.

Случайная переменная ξ имеет нормальное распределение, если функцией ее распределения является следующая функция:

$$P(\xi < x) = \frac{1}{\sqrt{2\pi\sigma}} \int_{-\infty}^{x} e^{-\frac{(t-m)^2}{2\sigma}}\, dt,$$

где m — любое вещественное, σ — любое положительное число.

В таком случае функция плотности величины ξ:

$$\boxed{f(x) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(x-m)^2}{2\sigma^2}},}$$

математическое ожидание: $M(\xi) = m$, дисперсия: $D(\xi) = \sigma$.

Функция нормального распределения с параметрами 0,1 уже была представлена в связи с формулой Муавра—Лапласа.

ПРАВИЛА БОЛЬШИХ ЧИСЕЛ

Наглядное понятие вероятности связано с понятием относительной частоты. При математической формулировке теории вероятности данное наглядное представление не нашло применения. Но если представленная теория соответствует нашим соображениям о понятии вероятности, то следует доказать, что ожидаемые соотношения на самом деле действительны между вероятностью и относительной частотой. Такие соотношения представлены в правилах больших чисел.

Первый результат такого рода уже присутствует в книге Я. Бернулли (1713 г.).

Пусть A представляет собой некоторое событие с вероятностью p. Пусть проводится n независимых экспериментов, результатом каждого из которых является наступление или ненаступление события A; пусть k означает частоту события A в n экспериментах. Тогда, если допустить какое угодно малое отклонение между вероятностью и относительной частотой, допустимое отклонение не будет превышено с какой угодно большой вероятностью, если количество экспериментов является довольно большим. Данное утверждение выражается в математической форме следующим образом. Если ε и δ представляет собой какие угодно маленькие положительные числа, и значение n_0 достаточно велико, то вероятность того, что относительная частота отличается от вероятности на величину больше ε будет меньше, чем δ, если количество экспериментов больше, чем n_0. В виде формулы:

$$P\left\{\left|\frac{k}{n} - p\right| > \varepsilon\right\} < \delta, \qquad \text{если} \qquad n \geqq n_0.$$

Здесь значение n_0 зависит от ε, δ и, естественно, от числа p.

В правильности данной теоремы можно убедиться следующим образом.

Утверждение, по которому относительная частота отличается от вероятности в крайнем случае на величину ε,

$$\left|\frac{k}{n} - p\right| \leqq \varepsilon,$$

может быть сформулировано и таким образом, что частота находится между значениями $np - p\varepsilon$ и $np + n\varepsilon$, т. е.

$$np - n\varepsilon \leqq k \leqq np + n\varepsilon.$$

Вероятность такого события составляет

$$\sum_{k=np-n\varepsilon}^{np+n\varepsilon} \binom{n}{k} p^k(1-p)^{n-k}.$$

Если применять формулу Муавра—Лапласа, то получится следующее соотношение:

$$\sum_{k=np-n\varepsilon}^{np+n} \binom{n}{k} p^k(1-p)^{n-k} \sim \Phi\left(\frac{n\varepsilon}{\sqrt{np(1-p)}}\right) - \Phi\left(-\frac{n\varepsilon}{\sqrt{np(1-p)}}\right) =$$

$$= 2\Phi\left(\frac{\varepsilon}{p(1-p)}\sqrt{n}\right) - 1.$$

Следовательно, необходимо показать, что число n_0 может быть выбрано таким образом, чтобы удовлетворилось неравенство

$$2\Phi\left(\frac{\varepsilon}{\sqrt{p(1-p)}}\sqrt{n}\right) - 1 \geqq 1 - \delta.$$

(Так как в таком случае вероятность обратного события $\left(\left|\frac{k}{n} - p\right| \geqq \varepsilon\right)$ меньше, чем δ.) То есть, необходимо показать, что число n может быть выбрано таким образом, чтобы удовлетворилось следующее неравенство:

$$\Phi\left(\frac{\varepsilon}{\sqrt{p(1-p)}}\sqrt{n}\right) \geqq 1 - \frac{\delta}{2}.$$

Поскольку функция Φ является монотонно нарастающей и приближается к числу 1, то приведенное выше неравенство выполняется при выборе достаточно большого значения n.

Известны различные обобщения данной теоремы. В обобщениях в основном представлены такие результаты, по которым утверждается, что при большом количестве экспериментов среднее число результатов многих независимых экспериментов (случайной переменной) будет близко к среднему значению математического ожидания данных случайных переменных.

ТЕОРЕМА О ЦЕНТРАЛЬНОМ ГРАНИЧНОМ РАСПРЕДЕЛЕНИИ

В теории вероятности одной из важных проблем является исследование вопроса о том, какие результаты получаются при сложении большого числа случайных независимых переменных. (Правила больших чисел также дают ответ на этот вопрос.) Между прочим, данный вопрос является существенным, так как некоторое случайное воздействие часто образуется в результате сложения большого числа независимых случайных воздействий. (Например, формирование средней температуры одного дня можно считать результатом сложения очень многих независимых случайных явлений.)

В теоремах о центральном граничном распределении утверждается, что сложение большого числа независимых случайных переменных имеет нормальное распределение при весьма общих условиях. Этим и объясняется, почему так часто встречается на практике нормальное распределение.

Одной из самых простых теорем о центральном граничном распределении является следующая теорема.

Пусть $\xi_1, \xi_2, \ldots$ случайные переменные с одинаковыми распределениями с ограниченным математическим ожиданием m и дисперсией σ. Тогда действительно следующее соотношение:

$$\boxed{P\left(\frac{\xi_1+\xi_2+\ldots+\xi_n-nm}{\sqrt{n}\sigma}<x\right)\to\Phi(x)\quad(n\to\infty).}$$

Заметим, что значение nm в формуле является математическим ожиданием суммы $\xi_1+\xi_1+\ldots+\xi_n$, а значение $\sqrt{n}\sigma$ — дисперсией суммы. Таким образом, математическое ожидание и дисперсия случайной переменной

$$\boxed{\frac{\xi_1+\xi_2+\ldots+\xi_n-nm}{\sqrt{n}\sigma}}$$

будут 0 и 1 соответственно.

Представленная теорема о центральном граничном распределении включает в себя теорему Муавра—Лапласа. Пусть $\xi_1, \xi_2 \ldots, \xi_n$ — независимые случайные переменные

$$P(\xi_i = 1) = p; \qquad P(\xi_i = 0) = 1 - p \quad (i = 1, 2, \ldots)$$

с общим распределением. Тогда очевидно, что

$$P(\xi_1 + \xi_2 + \ldots + \xi_n = k) = \binom{n}{k} p^k(1-p)^{n-k} \; (k = 0, 1, 2, \ldots, n;\; n = 1, 2, \ldots)$$

и

$$\sum_{k=l}^{j} \binom{n}{k} p^k(1-p)^{n-k} = P(l \leqq \xi_1 + \xi_2 + \ldots + \xi_n \leqq j) =$$

$$= P\left(\frac{l-np}{np(1-p)} = \frac{\xi_1 + \xi_2 + \ldots + \xi_n - np}{np(1-p)} \leqq \frac{j-np}{np(1-p)}\right).$$

Таким образом, из теоремы о центральном граничном распределении получится формула Муавра—Лапласа.

СТОХАСТИЧЕСКИЕ ПРОЦЕССЫ

Часто ставятся такие эксперименты, результаты которых нельзя охарактеризовать единственным числом, а только функцией (пример 4 на стр. 584). Глава теории вероятности, занимающаяся проблемами случайных функций, носит название теории стохастических процессов. Термин «процесс» здесь указывает на то, что в общем случае встречаются такие случайные функции, независимой переменной которых является время. Для стохастических процессов обычно применяется обозначение ξ_t, что означает, что в любой фиксированный момент времени t величина процесса ξ_t является числом, зависящим от случайности (случайная переменная), в то время как путем изменения величин получится случайная функция.

Как правило, стохастические процессы разделяются в зависимости от того, какие связи существуют между случайными переменными ξ_{t_1}, $\xi_{t_2}, \ldots, \xi_{t_n}$, соответствующими моментам времени $t_1, t_2 \ldots, t_n$. Для примера здесь приводится единственный тип стохастического процесса — стационарный процесс. Например, число бракованных деталей, производимых некоторой машиной, представляет собой стационарный процесс, если в данный промежуток времени качество машины не ухудшается и рабочий, работающий на машине, не устает (не ослабляется его внимание). Если в рассматриваемый интервал времени следует учитывать и износ машины, то число бракованных деталей обычно увеличивается (т. е. математическое ожидание количества бракованных деталей растет) и процесс является нестационарным.

Точнее, стохастический процесс ξ_t называется стационарным, если выполняется соотношение

$$P(\xi_{t_1} < x_1, \xi_{t_2} < x_2, \ldots, \xi_{t_n} < x_n) = P(\xi_{t_1+\tau} < x_1, \xi_{t_2+\tau} < x_2, \ldots, \xi_{t_n+\tau} < x_n)$$

для любых значений $t_1, t_2, \ldots, t_n, x_1, x_2, \ldots, x_n, \tau$.

Исследование стационарных процессов является особенно важным в области связи. Это очевидно, так как шум, мешающий передаче, представляет собой стохастический процесс, который, можно полагать, является стационарным процессом. При фильтрации шума большое значение имеет знание свойств стационарных процессов. Можно сделать вывод, что вообще шум является так называемым гауссовским процессом, означающим, что ξ_t (для любого значения t) представляет собой случайную переменную с нормальным распределением, и процесс имеет независимые разности, т. е., если удовлетворяются неравенства $t_1 < t_2 \leq t_3 < t_4 \leq \leq \cdots \leq t_{2n-1} < t_{2n}$, то случайные переменные $\xi_{t_2} - \xi_{t_1}$, $\xi_{t_4} - \xi_{t_3}, \ldots$, $\xi_{t_{2n}} - \xi_{t_{2n-1}}$ являются независимыми случайными величинами.

ТЕОРИЯ ИНФОРМАЦИИ

Основная проблема теории информации и необходимость математического измерения количества информации возникли в процессе изучения проблем связи. Мера для количества информации была введена американским математиком Шенноном, работавшим в лаборатории по телефонным связям фирмы Белл, и исследование таких проблем стало необходимым в ходе работ, проводимых его лабораторией.

Предположим, что какое-либо сообщение следует передавать таким образом, что в ходе передачи сообщения применяются только два знака (например, 0 и 1). Такое положение существует, например, в телеграфных устройствах системы Морзе. Предположим далее, что данное сообщение представляет собой одно из предварительно зафиксированных n возможных сообщений. Такое положение возникает при передаче букв ($n = 30$ по английскому алфавиту), или если, например, передается средняя температура дня, то сообщение представляет собой цифру от -20 до $+30$ ($n = 51$).

Ставится вопрос: как может быть установлено соответствие прогрессии знаков 0 и 1 к каждому из возможных сообщений таким образом, чтобы по возможности короткие прогрессии соответствовали отдельным сообщениям.

Пусть будут, например, возможные сообщения a, b, c, d, e, f, g, h ($n = 8$); надо установить соответствие символов, состоящих из трех цифр 0, 1, каждому сообщению таким образом, чтобы различным сообщениям соответствовали различные прогрессии. Это осуществляется следующим образом:

$$\begin{array}{ll} a \to 000 & e \to 100 \\ b \to 001 & f \to 101 \\ c \to 010 & g \to 110 \\ d \to 011 & d \to 111. \end{array}$$

В общем случае, если $n = 2^k$, то с помощью прогрессий длины k сообщение может быть передано, так как каждому сообщению можно предписать прогрессию, состоящую из 0 и 1, с длиной k таким образом, чтобы любым двум сообщениям соответствовали различные прогрессии знаков.

Можно показать, что для любого числа n сообщение можно передать с помощью прогрессий длины в среднем $\overset{2}{\log}\, n$.

Если известно, что n возможных сообщений поступают с неодинаковой вероятностью, то имеется возможность найти такое «кодирование» (отдельным сообщениям предписать различные прогрессии знаков), чтобы в среднем получилась прогрессия знаков короче, чем $\overset{2}{\log}\, n$. Например, при передаче средних температур дня выгодно такое кодирование, при котором часто встречающимся значениям температуры предписывается короткая прогрессия знаков, даже и в том случае, если в результате этого редко встречающимся значениям температуры предписываются более длиные прогрессии знаков.

Формула Шеннона. Предположим, что n возможных сообщений наступают с вероятностью

$$p_1, p_2, \ldots, p_n \qquad (p_1 + p_2 + \ldots + p_n = 1)$$

соответственно. В таком случае может быть показано, что можно найти такое кодирование, с помощью которого математическим ожиданием прогрессий знаков является следующее выражение:

$$\boxed{\sum_{k=1}^{n} p_k \overset{2}{\log} \frac{1}{p_k}\,.}$$

Лучше данного кодирования найти нельзя. Это и есть формула Шеннона. (Точнее, и к данному значению можно приблизиться только в том случае, если передается более длинное сообщение, буквами которого являются члены n возможных сообщений.)

В области техники связи передаче сообщения препятствует шум. Поэтому в теории информации рассматривается и следующий вопрос: найти такое кодирование, при котором задача декодирования может быть решена (может разгадываться, какое сообщение было передано), даже в том случае, если несколько знаков неясны из-за присутствия шума.

О НЕКОТОРЫХ ДРУГИХ ПРОБЛЕМАХ ТЕОРИИ ВЕРОЯТНОСТИ

Новейшие проблемы развития и результаты теории вероятности, по-видимому, наиболее целесообразно рассматривать на примерах их практического применения, тем более, что в последние годы во многих областях науки и техники открылись большие возможности применения методов теории вероятности. Приведем некоторые примеры.

а) *Биология.* В области биологических, медицинских и сельскохозяйственных исследований играют важную роль прежде всего статистические методы (ниже, при обсуждении статистических методов мы еще вернемся к данному вопросу). Из возможных областей применения теории вероятности здесь упоминается только о генетике.

Самая простая модель генетики имеет следующий вид: каждый ген имеет две возможные формы: A и a. В каждой клетке присутствуют два гена, соответствующие некоторому данному свойству (за исключением половых клеток). Таким образом, три различные пары (генотипа) встречаются в отдельных клетках: AA, Aa, aa (между вариантами Aa и aA нет разницы). Каждый генотип соответствует одному наследственному свойству. В половых клетках каждому свойству соответствует только один ген. В половой клетке индивида генотипа Aa форма гена A встречается с вероятностью $1/2$, и форма a — с вероятностью $1/2$, поэтому потомок индивидов генотипа Aa и генотипа Aa будет иметь генотип AA с вероятностью $1/4$, генотип aa — с вероятностью $1/4$ и генотип Aa — с вероятностью $1/2$.

На практике вопрос ставится следующим образом: если скрестить большое количество животных, имеющих определенные свойства, то потомки каких свойств ожидаются? (Решение такого типа задачи предполагает знание таких методов теории вероятности, представление которых требует более подробных рассуждений, превосхождящих возможности настоящей книги).

б) *Лингвистика.* В разных языках отдельные звуки воспроизводятся с различными частотами. Частота (вероятность) появления отдельных звуков в некоторой степени характерна для отдельных языков. Можно получить более точное представление, если одновременно рассматривать и вопрос, что *после* определенных звуков с какой частотой появляются другие звуки; или проводится подобное исследование в отношении слогов или слов. На основе применения такого метода исследований характерные данные могут быть получены не только в отношении отдельных языков, но и в отношении отдельных писателей. Таким образом, откроется возможность определения неизвестного ранее автора какого-либо произведения.

Аналогичного характера исследования могут быть проведены и в области музыки.

в) *Методы Монте-Карло.* Бывает, что расчет некоторых величин по точным математическим методам затрудняется, но может быть найдено такое событие, вероятность или математическое ожидание которого совпадает с искомой величиной. В таком случае значение искомой величины можно найти и таким образом, что определяется относительная частота данного события, и она считается приближенным значением величины. Таким образом, получится хорошее приближение с большой вероятностью.

На практике, вместо того, чтобы много раз проводить эксперимент, результат «случайных» экспериментов определяется с помощью электронных вычислительных машин из случайной прогрессии («орел или решка»). Для того, чтобы машина могла выполнить такую работу, необходимо подготовить для нее случайную прогрессию «орел или решка». На первый взгляд это, кажется, просто, но на деле — вопрос, является ли она «достаточно случайной», — весьма сложный. Эти проблемы разрабатывал А. Н. Колмогоров. Его результаты оказались настолько эффективными, что по-видимому стало необходимым новое аксиоматическое построение теории вероятности.

МАТЕМАТИЧЕСКАЯ СТАТИСТИКА

ЗАДАЧА МАТЕМАТИЧЕСКОЙ СТАТИСТИКИ

В теории вероятности обычно исходят из того, что известны или вероятность некоторых более простых событий, или функции распределения некоторых случайных переменных. При подбрасывании монеты, например, нам известно (как предварительное условие), что подбрасываемая монета имеет симметричную форму. Но это не доказано нами. Может возникнуть вопрос, имеет ли подбрасываемая монета действительно симметричную форму. Симметричность монеты может быть проверена применением некоторого физического метода. Если данная монета подбрасывается много раз подряд и орел выпадает примерно в половине всех экспериментов, то можно предположить, что монета является симметричной. Чтобы высказать такое предположение, необходимо определить, что подразумевается под термином «монета подбрасывается *много* раз» и «*примерно* в половине». С помощью математической статистики можно ответить на эти вопросы, так как ее задачей является определение на основе результатов экспериментов вероятности некоторых событий, функции плотности случайных переменных, математического ожидания, дисперсии и других характерных черт. Но бывает (небольшая вероятность), что монета выпадет на орел много раз подряд ; в таком случае принимается решение : монета не имеет симметричную форму. Но может случиться, что монета несимметричной формы выпадает на орел именно в половине первой части экспериментов и в этом случае также принимается ошибочное решение. На основе применения методов математической статистики можно утверждать о правильности решения с большой вероятностью.

ИССЛЕДОВАНИЕ ГИПОТЕЗ

Часто приходится ответить на вопрос о правильности или неправильности некоторой гипотезы. Например, если на каком-то заводе приняли определенную партию товара, то не обязательно знать точно, сколько процентов составляют бракованные предметы в партии товара. Обычно достаточно определить, что относительное количество бракованного товара больше или меньше, чем 3%. В первом случае партия товара будет отвергнута, а во втором — принята. Для реализации исследований статистических гипотез служат статистические пробы.

Проба u. Пусть ξ представляет собой случайную переменную с нормальным распределением, с известной дисперсией σ. Надо проверить такую гипотезу, равняется ли математическое ожидание случайной величины ξ определенному числу m_0.

Для контроля данной гипотезы выбирается проба из n элементов таким образом, что элементы пробы выбираются независимо друг от друга. Пусть результатом снятия проб будут величины $\xi_1, \xi_2, \ldots, \xi_n$. Проведем расчет величины

$$u = \frac{\xi_1 + \xi_2 + \ldots + \xi_n - nm_0}{\sqrt{n}\,\sigma}.$$

Можно увидеть, что случайная переменная u имеет нормальное распределение, дисперсию 1; если выполняется гипотеза $M(\xi) = m_0$, то $M(u) = 0$. Таким образом, при выполнении данной гипотезы действительно следующее соотношение:

$$P(|u| \leq \varepsilon) = \frac{1}{\sqrt{2\pi}} \int_{-\varepsilon}^{+\varepsilon} e^{-\frac{t^2}{2}}\, dt = \Phi(\varepsilon) - \Phi(-\varepsilon) = 2\,\Phi(\varepsilon) - 1.$$

Пусть ε будет таким, что можно было выполнить соотношение

$$\frac{1}{\sqrt{2\pi}} \int_{-\varepsilon}^{+\varepsilon} e^{-\frac{t^2}{2}} dt = 0{,}95.$$

В таком случае, если из пробы получено такое значение u, абсолютная величина которого больше, чем ε, то гипотеза $M(\xi) = m_0$ может быть отвергнута, так как произошло событие, которое является невероятным при правильности гипотезы. Иными словами, если гипотеза $M(\xi) = m_0$ отвергнута при выполненнии условия $|u| > \varepsilon$, то только в 5% всех случаев будет отвергнута правильная гипотеза.

Бывает и так, что гипотеза $M(\xi) = m_0$ не выполняется и все-таки принимается решение о правильности гипотезы. Вероятность ошибки такого рода нельзя определить так просто, как в случае проявления ошибки другого характера. Ибо в этом случае, если $M(\xi) \neq m_0$, все равно $|M(\xi) = = m_0|$ представляет собой малую величину. В таком случае более вероятно, что мы примем неправильную гипотезу $M(\xi) = m_0$, чем в том случае, когда $|M(\xi) - m_0|$ имеет очень большое значение. В случае, если $M(\xi) = = m \neq m_0$ гипотеза принимается с вероятностью

$$P(m) = \sqrt{\frac{n}{2\pi}} \cdot \frac{1}{\sigma} \int_{m_0-\varepsilon}^{m_0+\varepsilon} e^{-\frac{(t-m)^2 n}{2\sigma^2}}\, dt.$$

Данная формула показывает, что если значение m близко к значению m_0, то с вероятностью ≈ 1 принимается гипотеза $M(\xi) = m_0$; если значение $|m - m_0|$ велико, то гипотеза отвергается с вероятностью ≈ 1.

Функция $P(m)$ называется силовой функцией пробы.

Проба t. При использовании пробы u незнание дисперсии обычно представляет собой серьезное затруднение. Пусть ξ означает случайную переменную с нормальным распределением. Проверяется гипотеза о том, равняется ли математическое ожидание случайной переменной ξ определенному числу m_0. Проведем n экспериментов, результаты которых обозначим знаками $\xi_1, \xi_2, \ldots, \xi_n$. Оценивается значение выражения

$$t = \frac{\sum_{k=1}^{n} \xi_k - nm}{\sigma_n^*},$$

где

$$\sigma_n^* = \sqrt{\sum_{k=1}^{n} (\xi_k - \bar{\xi})^2}$$

и

$$\bar{\xi} = \frac{\xi_1 + \xi_2 + \ldots + \xi_n}{n}.$$

Функция распределения случайной переменной t оценивается просто, и ее значения могут быть собраны в таблицах. Имея таблицы, можно решить вопрос о правильности гипотезы $M(\xi) = m_0$, как это было показано при представлении пробы u.

При применении подобных проб можно решить вопрос о принятии или отклонении других статистических гипотез. Ниже приведены только некоторые виды гипотез, для которых уже разработаны пробы.

Исследование независимости. Следует решить вопрос о гипотезе независимости двух случайных переменных.

Исследование совпадения. Решается гипотеза о совпадении функции распределения некоторой случайной переменной с определенной заданной функцией распределения.

Исследование однородности. Решается вопрос о гипотезе совпадения функций распределений двух случайных переменных.

ТЕОРИЯ ОЦЕНКИ

Часто встречаются такие вопросы, на которые нельзя просто ответить «да» или «нет». Например, недостаточно определить, каков процент бракованных предметов в некотором объеме товара ниже определенного уровня, но и необходимо дать оценку значения процентов бракованных предметов. Задача теории оценки заключается в оценке функции распределения заданной случайной переменной или некоторых ее параметров (математического ожидания, дисперсии) на основе пробы.

ОЦЕНКА МАТЕМАТИЧЕСКОГО ОЖИДАНИЯ

Ставится задача оценки неизвестного значения математического ожидания m некоторой случайной переменной ξ из независимых элементов пробы $\xi_1, \xi_2, \ldots, \xi_n$. Например, ξ означает рост венгерского мужчины, выбранного случайно, а $\xi_1, \xi_2, \ldots, \xi_n$ — рост n мужчин, выбранных случайно. Кажется естественным применять число

$$\bar{\xi} = \frac{\xi_1 + \xi_2 + \ldots + \xi_n}{n}$$

для оценки значения величины m. Отметим, что по закону больших чисел величина $\bar{\xi}$ будет близка к математическому ожиданию m при большом количестве пробы. Хотя данный факт показывает, что величина $\bar{\xi}$ действительно является некоторой оценкой математического ожидания m, тем не менее в определенных случаях его нельзя считать лучшей оценкой из возможных величины m. Элементы пробы $\xi_1, \xi_2, \ldots, \xi_n$ упорядочиваются по их величине; полученная таким образом упорядоченная проба обозначается знаками $\xi_1^* \leq \xi_2^* \leq \ldots \leq \xi_n^*$. Бывают случаи, когда величина

$$\frac{\xi_1^* + \xi_2^*}{2}$$

в определенном смысле является более правильной оценкой значения неизвестного математического ожидания.

ОЦЕНКА ДИСПЕРСИИ

Самой естественной оценкой неизвестного значения дисперсии является величина

$$\sqrt{\frac{\sum_{k=1}^{n}(\xi_k - \bar{\xi})^2}{n}},$$

при которой (также на основе законов больших чисел) можно увидеть, что ее значение близко к значению дисперсии при большом n.

НЕКОТОРЫЕ ОБЩИЕ ПОНЯТИЯ ТЕОРИИ ОЦЕНКИ

Ставится задача определения оценки неизвестного параметра a функции распределения случайной переменной ξ с помощью элементов пробы $\xi_1, \xi_2, \ldots, \xi_n$.

Некоторая функция $f_n(\xi_1, \xi_2, \ldots, \xi_n)$ переменных $\xi_1, \xi_2 \ldots, \xi_n$ называется статистикой. Статистика f_n называется оценкой параметра без искажения, если выполняется соотношение

$$M(f_n(\xi_1, \xi_2, \ldots, \xi_n)) = a .$$

Прогрессия статистики $f_1, f_2, \ldots$ называется консистенциальной оценкой параметра a, если значение статистики $f_n(\xi_1, \xi_2, \ldots, \xi_n)$ близко к значению параметра a с большой вероятностью при больших n, точнее, если для каждого положительного значения ε и δ может быть найдено такое целое число n_0, при котором удовлетворяется неравенство

$$P(|f_n(\xi_1, \xi_2, \ldots, \xi_n) - a| \geqq \varepsilon) \leqq \delta ,$$

если $n \geqq n_0$.

Утверждение, по которому некоторая функция рассматривается ст элементов пробы $\xi_1, \xi_2, \ldots, \xi_n$ при оценке некоторого параметра, т. е. рассматривается единственное число, часто означает, что теряется некоторая информация, присутствующая в пробе a, относящаяся к параметру a. В случае, когда этого не происходит, функция f называется *достаточной статистикой*. (Здесь не приводится точное математическое определение понятия достаточной статистики.)

ОЦЕНКА ФУНКЦИЙ РАСПРЕДЕЛЕНИЯ И ПЛОТНОСТИ

В случае, когда нет никакой информации о функции распределения некоторой случайной переменной, приходится оценивать полную функцию распределения из элементов пробы.

Пусть $\xi_1, \xi_2, \ldots, \xi_n$ представляет собой пробу из n элементов, а упорядоченная проба, полученная из данной пробы, обозначается символом $\xi_1^*, \xi_2^*, \ldots, \xi_n^*$. Функция распределения, основанная на опыте $F_n(x)$, определяется следующим образом:

$$F_n(x) = \begin{cases} 0, & \text{если} \quad x \leqq \xi_1^* \\ \dfrac{k}{n}, & \text{если} \quad \xi_k^* \leqq x < \xi_{k+1}^* \\ 1, & \text{если} \quad x \leqq \xi_n^* . \end{cases}$$

Поскольку функция $F_n(x)$ является относительной частотой такого события, при котором результат эксперимента меньше, чем x и функция $F(x)$ (неизвестная функция распределения) является вероятностью такого же события, то по законам больших чисел функция $F_n(x)$ будет близка к функции $F(x)$ с большой вероятностью при достаточно большом n. Это означает, что функцию $F_n(x)$ можно считать некоторой оценкой функции $F(x)$.

Гистограмма. Для оценки функции плотности применяется так называемая гистограмма. Гистограмма пробных элементов $\xi_1, \xi_2, \ldots, \xi_n$ может быть получена следующим образом. Предположим, что все полученные пробные элементы находятся между числами a и b, т. е.

$$a \leqq \xi_1^* < \xi_2^* < \ldots < \xi_n^* \leqq b .$$

Разделим интервал (a, b) на ν одинаковых частей с точками разделения $a = x_0 < x_1 < \ldots < x_{\nu-1} < x_\nu = b$ и рассчитаем относительную частоту такого события, при котором некоторый пробный элемент попадает между точками x_k и x_{k+1}. Если данную относительную частоту изобразить вертикально над интервалом $(x_k; x_{k+1})$, то получится *гистограмма*. Путем соответствующего выбора значения ν (следует выбирать примерно $\sqrt{n}$) может быть достигнуто состояние, при котором гистограмма будет близка к функции плотности с большой вероятностью.

ИНТЕРВАЛ КОНФИДЕНЦИИ

Лишь в редких случаях можно довольствоваться тем, что найдена некоторая оценка неизвестного параметра: обычно приходится определить, насколько данная оценка является точной, иными словами

следует найти такой интервал, содержащий неизвестный параметр с большой вероятностью. Такой интервал называется интервалом конфиденции.

Метод определения интервала конфиденции можно показать на следующем примере. Пусть ξ представляет собой случайную переменную с нормальным распределением, с известной дисперсией σ и с неизвестным значением математического ожидания m. Пусть $\xi_1, \xi_2, \ldots, \xi_n$ представляют собой такую пробу, из которой следует определить интервал конфиденции. Как известно, значение математического ожидания m оценивается с помощью статистики

$$\bar{\xi} = \frac{\xi_1 + \xi_2 + \ldots + \xi_n}{n},$$

поэтому естественно искать интервал конфиденции в виде $(\bar{\xi} - a, \bar{\xi} + a)$. Точнее, задача заключается в таком определении значения a, чтобы интервал $(\bar{\xi} - a, \bar{\xi} + a)$ содержал в себе неизвестный параметр m с большой вероятностью. Нетрудно увидеть, что

$$P(\bar{\xi} - a < m < \bar{\xi} + a) = \frac{\sqrt{n}}{\sqrt{2\pi}\,\sigma} \int_{-a}^{a} e^{-\frac{t^2 n}{2\sigma^2}} dt .$$

Таким образом, с помощью таблицы, представленной для функции Φ, может быть определено значение a при условии, что интервал $(\bar{\xi} - a;\ \bar{\xi} + a)$ содержит в себе значение m с желаемой вероятностью.

ТЕОРИЯ ИГР

Основателем теории игр считают математика Иоганна фон Неймана. Им доказаны основная теорема теории игр и возможности ее практического применения.

В теории игр обычно речь идет об играх двух лиц; предполагается, что игры являются ограниченными, т. е. оба игрока в обоих случаях могут делать только определенное количество шагов и игра заканчивается после ограниченного количества шагов. Отсюда вытекает ограниченность числа стратегий обоих игроков.

Под понятием стратегии подразумевается такая система правил, с помощью которой задается, как ходит игрок в некоторой данной ситуации.

В игре бывают и случайные ходы (например, подбрасывание игральной кости, раздача карт и т. д.). Таким образом, даже знанием стратегий обоих игроков не определяется однозначно исход игры. Пусть $M(i, j) = a_{ij}$ означает значение математического ожидания выигрыша первого игрока при условии, что первый игрок проводит игру по i-той, а второй — по j-той стратегии. В соответствии с этим вся игра может рассматриваться с помощью матрицы размерности $n \times m$, где n означает число стратегий первого, а m — второго игрока. Элемент матрицы a_{it} в i-той строке и в j-том столбце представляет собой значение математического ожидания первого игрока. Такая игра иногда называется матричной игрой.

Стратегия. Целью теории игр (обычно в матричных играх) является нахождение лучшей стратегии отдельных игроков. В большинстве игр было бы нецелесообразно применять всегда одну и ту же стратегию, так как если противник определит нашу стратегию, ему легче будет найти хорошую обратную стратегию. Таким образом, кажется вероятным получение лучшей стратегии в том случае, когда стратегия выбирается случайным образом из отдельных стратегий. Предположим, что первый игрок выбирает свою стратегию из возможных с вероятностью $p_1, p_2, \ldots, p_n$ $(p_1 + p_2 + \ldots + p_n = 1)$, а второй — с вероятностью $q_1, q_2, \ldots, q_m$ $(q_1 + q_2 + \ldots + q_m = 1)$, соответственно. В таком случае математическое ожидание выигрыша отдельных игроков определяется выражением:

$$\sum_{i=1}^{n} \sum_{j=1}^{m} a_{ij} p_i q_j .$$

Цена игры. Первый игрок может проводить следующее рассуждение: если я выбираю свои стратегии с вероятностями $p_1, p_2, \ldots, p_n$, то могу предположить, что мой противник будет выбирать свои с вероятностями $q_1, q_2, \ldots, q_m$, при которых математическое ожидание моего выигрыша будет очень мало. Иначе говоря, если я буду выбирать из моих стратегий по распределению вероятностей $p_1, p_2, \ldots, p_n$, то математическое ожидание моего выигрыша будет, по меньшей мере,

$$\inf_{\{q\}} \sum_{i=1}^{n} \sum_{j=1}^{m} a_{ij} p_i q_j .$$

где инфимум относится ко всем распределениям вероятностей $q_1, q_2, \ldots, q_m$. Итак, лучшим можно считать такое распределение вероятностей $p_1, p_2, \ldots, p_n$, для которого самое большое возможное значение имеет следующее выражение:

$$\inf_{\{q\}} \sum_{i=1}^{n} \sum_{j=1}^{m} a_{ij} p_i q_j .$$

Аналогично, для второго игрока лучшей стратегией является выбор по распределению вероятностей $q_1, q_2, \ldots, q_m$, для которого значение выражения

$$\sup_{\{p\}} \sum_{i=1}^{n} \sum_{j=1}^{m} a_{ij}p_iq_j$$

имеет наименьшее возможное значение.

По основной теореме теории игр выполняется следующее соотношение:

$$\boxed{\sup_{\{p\}} \inf_{\{q\}} \sum_{i=1}^{n} \sum_{j=1}^{m} a_{ij}p_iq_j = \inf_{\{p\}} \sup_{\{q\}} \sum_{i=1}^{n} \sum_{j=1}^{m} a_{ij}p_iq_j = \lambda\,.}$$

Данное число называется ценой игры.

Итак, по основной теореме теории игр утверждается, что первый игрок имеет такую стратегию, при использовании которой математическое ожидание его выигрыша будет по меньшей мере λ (при выборе вторым игроком любой стратегии); а второй игрок имеет такую стратегию, при которой математическое ожидание его потери будет по крайней мере λ.

РЕШАЮЩИЕ ФУНКЦИИ

Теория решающих функций является одной из наиболее новых областей статистики. Хотя они применяются широко, решающие функции здесь представлены только на примере контроля качества товара. Очевидно, что в некотором объеме товара процент брака оцениваются тем точнее, чем больше элементов выбирается для пробы. Снятие пробы со многими элементами, однако, может быть связано с большими расходами. Может быть дорогостоящим и исследование элементов пробы, и бывает, когда в результате исследования элементы придут в негодное состояние. (Если, например, исследуется срок перегорания электрических ламп, приходится прокалить несколько электрических ламп.)

Так как снятие проб связано с расходами, статистику нужно стремиться принять решение на основе пробы с небольшим количеством элементов. Однако решение, принятое на основе пробы с небольшим количеством элементов, будет ошибочно с большей вероятностью, чем решение, принятое на основе пробы с большим количеством элементов. Ошибочное решение может привести к крупным потерям. Статистик должен выбрать такое число элементов пробы, при котором расходы (расходы снятия пробы плюс расходы из-за ошибочных решений) будут минимальными.

Пусть ξ — некоторая случайная переменная с известной $F(x, a)$ функцией распределения, с неизвестным параметром a. Например, ξ может означать некоторую характеристику поставляемого товара (например, вес, длину), и параметр a, например, может представлять математическое ожидание случайной величины ξ.

Решение следует принимать на основе пробы с n элементами $\xi_1, \xi_1, \ldots, \xi_n$. Предположим, что возможные решения могут быть $d_1, d_2, \ldots, d_r$. Решение принимается на основе исследования пробы $\xi_1, \xi_2, \ldots, \xi_n$, т. е. дается объяснение n-мерной *решающей функции* $d(\xi_1, \xi_2, \ldots, \xi_n)$, возможными значениями которой являются возможные решения $d_1, d_2, \ldots, d_r$.

Пусть обозначаются суммарные расходы через $L(a_0, d_i)$, если предположить, что значением параметра a является a_0, и в результате исследования пробы $\xi_1, \xi_2, \ldots, \xi_n$ принято решение d_i. Зная функции $L(a, d_i) = L(a, d(\xi_1, \xi_2, \ldots \xi_n))$ и $F(x, a)$ (функция распределения), математическое ожидание расходов станет исчислимым при условии $a = a_0$. Пусть такое математическое ожидание обозначается через

$$M(L(a_0, d(\xi_1, \xi_2, \ldots, \xi_n))) = m(a_0)\,.$$

Количество элементов пробы n и решающую функцию $d(\xi_1, \xi_2, \ldots, \xi_n)$ следует выбирать таким образом, чтобы значение $m(a_0)$ было по возможности мало при любом значении a_0, или иными словами, ищутся такое значение n и такая функция d, для которых значение $\sup\limits_{a_0} m(a_0)$ является минимальным.

Такой вывод делается на основе рассуждений, проводимых по теории игр. Представим себе, что статистик и «природа» играют; в этой игре природа имеет право выбирать значение параметра a, а правом статистика является выбирать решающую функцию. Иначе говоря, стратегиями природы являются возможные значения параметра a, а стратегиями статистика — возможные решения. (Такая игра обычно не является конечной.) По рассуждениям, проводимым в соответствии с теорией игр, очевидно, что самая хорошая стратегия статистика будет такая, что он случайно выбирает решение из возможных решений с некоторым распределением вероятностей.

Другой результат будет получен, если параметр a рассматривается как случайная переменная. Тогда ситуация рассматривается таким образом, что значение a является числом, зависящим от различных случайных факторов в некотором объеме товара, доставляемого в разных случаях. Предположим, что известна функция распределения $g(x)$ случайной переменной a. Тогда величина a также представляет собой случайную переменную; в таком случае статистический метод целесообразно выбирать таким образом, чтобы значение математического ожидания $m(a)$ было минимальным.

О НЕКОТОРЫХ ДРУГИХ ПРОБЛЕМАХ МАТЕМАТИЧЕСКОЙ СТАТИСТИКИ

Из многочисленных приложений математической статистики, применяемых как в различных научных исследованиях, так и на практике, здесь останавливаемся на единственной области применения — в исследованиях в области сельского хозяйства.

Предположим, надо получить новый сорт растения или разработать новый метод обработки (или нескольких методов обработки одновременно).

В таком случае на разных участках применяются различные методы обработки и рассматривается вопрос: какой метод обработки дает большую урожайность. В то же время возникает вопрос: если в результате одного из новых методов обработки получится лучший результат с небольшим отклонением, можем ли заключить, что данный метод обработки является лучшим, или нам придется предположить действие некоторых случайных факторов в незначительном повышении урожайности. Одна из задач статистиков заключается в определении такого уровня, при котором причиной повышенной урожайности считается лучший метод обработки. При планировании экспериментов в сельском хозяйстве большую трудность представляет тот факт, что необходимо считаться и с неоднородностью почвы. Поэтому необходимо повторять отдельные эксперименты на различных территориях и при этом следить за тем, чтобы без особого основания не использовать слишком большие участки и для экспериментальных целей. Оптимальное планирование экспериментов, для которого с математической точки зрения комбинаторные и алгебраические методы применяются в первую очередь, также является задачей статистиков.

КРАТКИЙ ОБЗОР ИСТОРИИ ТЕОРИИ ВЕРОЯТНОСТИ

Теория вероятности появилась в XV веке, в эпоху Возрождения, когда азартные игры — игральная кость, карты — были особенно распространены. Первыми выдающимися учеными в этой области науки были: Л. Пачоли (1445—1510), Н. Тарталья (1499—1557) и Дж. Кардано (1501—1576). Даже такие крупные ученые как Г. Галилей (1564—1642), Б. Паскаль (1623—1662), П. Ферма (1601—1665) и Х. Хейгенс (1629—1695) занимались проблемами азартных игр. Развитие торговли, возросшее значения выигрыша и риска в торговле, страхование, возникшее в связи с развитием судоходства, также способствовали созданию теории вероятности.

Важным этапом истории теории вероятности является посмертное издание книги „Ars conjectandi" (Искусство догадок) Я. Бернулли (1654—1705). В этой книге дается общее представление о теории вероятности, открывается закон о больших числах и обосновывается математическая теория вероятности. Позже Муавром (1667—1751), Байесом (1702—1761) и Лапласом (1749—1827) были достигнуты такие результаты, которые широко применяются и в наши дни. Многое сделали в разработке теории вероятности такие ученые как К. Ф. Гаусс (1777—1855), А. М. Лежандр (1752—1833) и Д. Пуассон (1781—1840). Дальнейшее развитие теория вероятности получила только во второй половине XIX века в работах великих русских математиков П. Л. Чебышева (1821—1894), А. А. Маркова (1856—1922) и А. М. Ляпунова (1857—1918). Начало математической статистике было положено в Англии в начале XX века.

Долгое время теория вероятности не считалась равноценной другим областям математики, так как не были в достаточной степени выяснены ее принципиальные основы. Это обстоятельство тормозило развитие естественных наук, особенно развитие современной физики.

Огромный вклад в аксиоматическое построение теории вероятности был внесен советским математиком А. Н. Колмогоровым (год рождения 1902) в 1933 г. Его достижения в разработке теории вероятности сделали ее одной из важнейших глав математики, результаты которой применяются в различных областях математики и наоборот, многие выводы математической дисциплины находят применение в теории вероятности.

За последние десятилетия круг применения теории вероятности значительно расширился в связи с развитием атомной физики, техники связи и автоматизации. Результаты теории вероятности широко применяются в решении как теоретических, так и практических задач, связанных с работой электронно-вычислительных машин.

В Венгрии выдающимся ученым в области теории вероятности и математической статистики являлся К. Йордан (1871—1959), а в последнее время А. Рени (1921—1970) и его ученики.

Советская школа по теории вероятностей продолжает занимать видное положение. Кроме упомянутого выше А. Н. Колмогорова среди советских ученых должны быть названы, прежде всего, имена С. Н. Бернштейна, А. Я. Хинчина и Б. В. Гнеденко.

В частности, в области математияеской статистики наиболее активными представителями советской школы являются А. Н. Ширяев, Р. Ш. Липцер и М. П. Ершов.

МАТЕМАТИЧЕСКАЯ ЛОГИКА

ПРЕДМЕТ МАТЕМАТИЧЕСКОЙ ЛОГИКИ

Математическая логика является современной формой так называемой *формальной логики*, применяющей математические методы для исследования своего предмета. (Другие ее названия: символическая логика, теоретическая логика, логистика.) В формальной логике и, соответственно, в математической логике, собраны результаты законов структуры правильных *выводов*. Вывод является таким мыслительным процессом, в результате которого появляются новые открытия на основании уже имеющихся (которые предполагаются правильными), без практических исследований. В действительности, новое открытие, полученное в результате вывода, (так называемый *окончательный вывод*) в скрытой форме находится в предварительно имеющихся знаниях, в так называемых *предпосылках*.

Простейшие закономерности выводов открывались человечеством *эмпирическим путем* в ходе общественного производства (например, простейшие соотношения арифметики и геометрии). Открытие более сложных законов связано с результатами науки формальной логики. Первое крупное обобщение формальной логики принадлежит Аристотелю. В формальной логике с самого начала применялись (в единичных случаях) математические методы, но развитие логики не успевало за применением таких методов по сравнению с другими областями математики. Поэтому формальная логика отстала от потребностей науки (в первую очередь от требований математики); отставание оказалось особенно очевидным в новую эру. Главными недостатками формальной логики являлись следующие:

1. Она не сумела привести законы выводов к небольшому количеству надежных логических законов; поэтому подтвердила правильность некоторых выводов на основе экспериментов, которые позже были опровергнуты примерами, доказывающими обратное.

2. Она была неспособна анализировать значительную часть выводов, применяемых в повседневной и научной жизни; доказать правильность или неправильность таких выводов. (Например, не могла доказать, что из правильности предложения «Каждая трапеция является четырехугольником» вытекает правильность предложения «Кто рисует трапецию, тот рисует четырехугольник»).

Задача математизации формальной логики была поставлена и осуществлена Лейбницем. Его работу продолжили математики XIX века. На рубеже столетия с открытием противоречий в теории множеств (см. гл. «Теория множеств») развитие математической логики получило широкий размах. В настоящее время результаты математической логики используются во всех традиционных областях формальной логики; открыты совершенно новые области. В настоящее время «традиционная» формальная логика по сравнению с математической логикой имеет значение только для истории науки.

Математическая логика не претендует на открытие законов мышления *вообще*, или еще в меньшей степени на анализ философских проблем, связанных с человеческим мышлением. Эти вопросы больше относятся к «логике» (в более общем смысле слова) и к философии. (В дальнейшем под словом «логика» будем подразумевать математическую логику.)

ЧТО ТАКОЕ ВЫВОД?

Для более точного определения предмета математической логики следовало бы уточнить, что подразумевается под термином логически правильного вывода. Чтобы сформулировать хотя бы одно временное определение, рассмотрим пример вывода. (В соответствии с традиционной формой записывания, предпосылки отделяются от окончательного вывода горизонтальной чертой):

(Предпосылки) Если будет раздача премии, то мы выполнили план.
Будет раздача премии.

(Окончательный вывод) Мы выполнили план.

Если принять правильность предпосылок, то следует принять и правильность окончательного вывода.

Другой, аналогичный пример:

Если мне выпадет туз, то я иду ва-банк.
Мне выпал туз.

Я иду ва-банк.

Обычно вместо предложений (мне выпал туз) и (я иду ва-банк) могут быть записаны любые такие изъявительные предложения, значения которых может быть правильно или ложно; следует оставить неизменными только расположение слов «если» и «то» и расположение предположений, то есть *структуру* вывода. Пусть A и B обозначает любые заменяющие предложения. *Структуру* вывода можно выразить следующей *схемой;*

$$\frac{\text{Если } A\text{, то } B \qquad A}{B}$$

Под определением, что данная схема представляет собой (логически правильную) *схему выводов*, подразумевается следующее. Если вместо A и B подставить такие предложения, что предпосылки, полученные в результате замены, будут правильными, то и окончательный вывод будет правильным. Любой человек, который понимает значение союзов «если . . . то», поймет, что это правильная схема вывода. В схеме вывода фигурируют несколько слов с *постоянным* значением, далее несколько символов (букв) с *меняющимся* значением. Символы с *меняющимся* значением могут быть переменными разных типов. В соответствии с их типом вместо символов могут быть подставлены разные грамматические формации (например: изъявительные предложения, слова, выражающие свойства, названия предметов и т. д.). В предыдущем примере переменные A и B заменяются только изъявительными предложениями. На основе «регулярной» замены переменных некоторой (правильной) схемы вывода должен возникать правильный вывод.

Но определение «регулярной замены» означает не только соблюдение грамматических правил. В предыдущей схеме A и B могут означать только такие изъявительные предложения, правильность или ложность которых может быть решена однозначно. Такие изъявительные предложения будем называть *высказываниями*.

На основе любой схемы вывода может быть получен правильный вывод только при соблюдении условий подобного характера. Путем изменения условий могут быть построены различные теории логики.

Важнейшими главами математической логики являются *калькуляция высказываний* и *калькуляция предикатов*. В рамках данных глав может быть исследована схема вывода в самом общем случае при наименьшем числе условий.

В других главах логики рассматриваются специальные схемы вывода, являющиеся менее общими.

КАЛЬКУЛЯЦИЯ ВЫСКАЗЫВАНИЙ

ВЫСКАЗЫВАНИЕ

Предметом калькуляции высказываний является анализ таких схем вывода, при которых с заменой переменных на высказывания, получаются правильные выводы.

Под термином высказывания подразумевается такое изъявительное предложение, которое является однозначно или правильным, или ложным; итак:

а) оно не может одновременно быть и правильным, и ложным (принцип непротиворечивости);

б) исключено, чтобы оно было и неправильным, и неложным (принцип исключения третьей возможности).

Свойства «правильное» и «ложное» подразумеваются в их обычном смысле; они не нуждаются в дальнейшем анализе.

При данных обстоятельствах приведенные выше изъявительные предложения удовлетворяют (с «хорошим приближением») этим двум условиям; их можно считать высказываниями. Поэтому логика, построенная на этих двух условиях, может получить весьма широкое применение. Естественно, существуют такие «тонкие обстоятельства», при которых некоторых изъявительных предложений нельзя считать высказываниями (например, если дано предложение: «Иван просыпается», вряд ли можно сомневаться в правильности или ложности предложения «Иван спит»). Математические термины определяются таким образом, что предложения, выражающие соотношения между ними, всегда считаются высказываниями; такое положение существует во всех точных науках.

Понятие «высказывание» иногда обозначается словами «утверждение», «суждение».

В выводах могут фигурировать высказывания (либо в виде предпосылок, либо как окончательный вывод), возникшие из одного или нескольких высказываний, путем применения некоторого грамматического метода; они называются *сложными высказываниями*. Во многих случаях правильность вывода зависит от вида формирования сложного высказывания. Поэтому необходимо заниматься видом формирования сложных высказываний некоторых типов.

Под термином калькуляции высказываний подразумевается такой метод, с помощью которого из одного или нескольких высказываний (членов операции калькуляции высказываний) получается такое высказывание (результат операции), правильность или ложность которого однозначно определяется правильностью или ложностью членов.

ОТРИЦАНИЕ И КОНЪЮНКЦИЯ

Двумя простейшими примерами вышеприведенной операции являются отрицание и конъюнкция. (Операция и результат операции здесь обозначается одним и тем же названием.)

Под отрицанием высказывания A подразумевается высказывание «Неправильно, что A» (или некоторая грамматически преобразованная форма данного высказывания).

По значению выражения «неправильно» отрицание A правильно тогда и только тогда, если самое A неправильно; следовательно, отрицание действительно есть операция калькуляции высказываний (в соответствии с вышеприведенным определением).

Пример: Отрицанием предложения «мотор работает» является предложение «неправда, что мотор работает» или, иначе: «мотор не работает».

Отрицание является *одночленной* операцией. Отрицание «A» обозначается символом «$\neg A$» (читается: «не A»). Применяются также и обозначения «$\sim A$», «$- A$», «$\overline{A}$».

Под конъюнкцией двух высказываний A и B подразумевается высказывание «A и B» (или некоторая грамматически измененная форма данного высказывания). По значению союза «и» конъюнкция является правильной тогда и только тогда, если оба ее члена правильны.

Таким образом, конъюнкция также является операцией калькуляции высказываний. Операция конъюнкции «A и B» представляет собой *дву-*

членную операцию; ее обозначают «$A \wedge B$», «A & B», «AB». При возникновении конъюнкции союз «и» иногда заменяется другим союзом (например, «Анатолий здесь, *но* Бориса нет» или «Анатолий здесь, *хотя* Борис ушел» и т. д.). Это не влияет на правильность или ложность результата, имеет только эмоциональное значение. Иногда союз вообще пропускается. Если сказуемые двух предложений, связанных между собой путем конъюнкции, совпадают, то общее сказуемое представлено только в одном из предложений. Например, конъюнкция «я питаюсь хлебом и питаюсь водой» после преобразования имеет следующий вид: «я питаюсь хлебом и водой».

Изучение остальных операций калькуляции высказываний уточняется и облегчается с помощью следующего рассуждения.

Пусть свойства высказываний «правильное» и «ложное» называются *логическими значениями* и обозначаются знаками n и $л$. Правильность (или ложность) некоторого высказывания A выражается и в такой форме, что логическим значением высказывания A является n (или $л$).

Если задаются логические значения отдельных членов в некоторой операции калькуляции высказываний, то данной операцией логическое значение результата определяется однозначно. Это позволяет определение таких операций для *логических значений* (кроме вышеприведенного определения для высказываний) следующим образом: На место и членов и результата подставляются логические значения; причем, вместо результата подставляется логическое значение высказывания, образующееся данной операцией из высказываний с соответствующими членам логическими значениями.

Например, отрицания логических значений определяются так:

$\neg n = л$ (так как отрицание правильного высказывания является ложным),

$\neg л = n$ (так как отрицание ложного высказывания является правильным);

а конъюнкции логических значений так:

$n \wedge n = n$ (так как конъюнкция двух правильных высказываний является правильной),

$$\left.\begin{aligned} n \wedge л &= л \\ л \wedge n &= л \\ л \wedge л &= л \end{aligned}\right\}$$ (так как если одно или оба из двух высказываний являются ложными, то и их конъюнкция будет ложной).

На основе вышеприведенного рассуждения изучение операций, проведенных на высказываниях, может быть заменено изучением операций, проведенных на логических значениях. Этого достаточно для исследования выводов (на уровне калькуляции высказываний).

АЛГЕБРА ЛОГИЧЕСКИХ ЗНАЧЕНИЙ

Операции, проводимые на логических значениях, называются *логическими операциями*. Для выражения *любых* логических значений вводятся *логические переменные*; они обозначаются символами p, q, r, . . ., p_1, p_2, . . . Итак, логические переменные могут принимать два «значения»: $п$ или $л$.

Представленные выше отрицание и конъюнкция определяются так:

$$\neg p = \begin{cases} л, \text{ если } p = п, \\ п, \text{ в любом другом случае}; \end{cases}$$

$$p \wedge q = \begin{cases} п, \text{ если } п = p = q, \\ л, \text{ в любом другом случае}. \end{cases}$$

При использовании нескольких операций последовательно порядок выполнения отдельных операций обозначается скобками; например, $\neg(p) \wedge q)$ (иногда скобки опускаются). Например, вместо выражения $(\neg p) \wedge q$ пишется $\neg p \wedge q$ при предварительном пояснении, что в случае появления выражения без скобок знак относится только к следующему знаку. Аналогично, вместо символа $\neg(\neg p)$ употребляется просто $\neg\neg p$.

На основе определений операций получены следующие соотношения:

$\neg\neg p = p$	(закон двойного отрицания),
$p \wedge p = p$	(идемпотенция конъюкции),
$p \wedge \neg p = л$	(закон противоречия),
$\neg(p \wedge \neg q) = п$	(закон непротиворечивности),
$p \wedge q = q \wedge p$	(коммутативность конъюнкции),
$(p \wedge q) \wedge r = p \wedge (q \wedge r)$	(ассоциативность конъюнкции).*

На основе последних двух тождеств повторная конъюнкция может быть записана без скобок, как «многочленная конъюнкция» (подобно тому, как это делается при сложении или произведении многих членов).

В общем смысле слова $п$-членной логической операцией называется каждая такая функция, областью существования которой является упорядоченное множество всех выражений, образуемых из логических значений $п$ и $л$ с длиной выражения $п$, а значением ее является одно из двух логических значений $п$ и $л$ (см. гл. «Теория множеств».)

* Читатель поступает правильно, если «переведет» данные соотношения (и следующие соотношения также) на язык операций, проводимых на высказываниях. Например, тождество «$q \wedge q = q \wedge p$» переводится следующим образом: «Если A и B представляют собой любые высказывания, то логическим значением выражения $A \wedge B$ является то же самое, что и логическое значение выражения $B \wedge A$».

Такую операцию можно представить в виде таблицы значений. (Для наглядного предварительного представления ниже дана таблица значений уже известных отрицания и конъюнкции):

p	$\neg p$
п	*л*
л	*п*

p	q	$p \wedge q$
п	*п*	*п*
п	*л*	*л*
л	*п*	*л*
л	*л*	*л*

Любая *n*-членная логическая операция, обозначаемая, например, в виде $f(p_1, p_2, \ldots, p_n)$, задается следующим образом: устанавливается, при каких значениях переменных $p_1, \ldots, p_n$ результат будет *п* и при каких — *л*. Простейшим, но не всегда кратчайшим методом такого представления является заполнение следующей *таблицы значений;*

p_1	p_2	$\ldots$	p_{n-1}	p_n	$f(p_1, \ldots, p_n)$
п	*п*	$\ldots$	*п*	*п*	
п	*п*	$\ldots$	*п*	*л*	

. . . .

В таблице значений имеется столько строк, сколько есть вариантов заполнения столбцов значениями *п* или *л*. Поскольку каждая строка может считаться вариацией с повторением *n*-го порядка из 2 элементов (из элементов *п* или *л*), то количество строк равняется числу таких вариаций, 2^n (см. гл. «Алгебра»)). Итак, *таблица значений n-членной логической операции* содержит 2^n *строк*. Но вместо полной таблицы достаточно задать те строки, в которых результатом операции является *п* (или — в которых стоит *л*). На основе вышеприведенного рассуждения можно сделать вывод, что последний столбец таблицы значений *n*-членной логической операции может быть заполнен по 2^{2^n} различным вариантам (так как в нем имеется 2^n мест). Следовательно, *число всех n-членных операций составляет* 2^{2^n}.

При $n = 1$ $2^{2^n} = 2^{2^1} = 4$, т. е. число одночленных операций составляет 4. При $n = 2$ $2^{2^n} = 2^{2^2} = 16$, т. е. количество двухчленных операций — 16. Количество трехчленных операций уже получится 256. Сформулируем теорему:

Любая логическая операция может быть выражена через операции отрицания и конъюнкции.

(Ниже будут приведены примеры применения этой теоремы).

В области логических операций для контроля любого тождества может быть применен следующий (не всегда самый короткий) метод: составим *общую* таблицу операций, представленных по обеим сторонам знака $=$.

В общую таблицу значений включаются все переменные обеих операций; результат операций указывается в столбцах, расположенных рядом друг с другом. Если два столбца содержат одинаковые результаты, то две операции являются тождественными.

Пример: Пусть контролируется правильность тождества

$$\neg(p \wedge q) \wedge \neg(p \wedge \neg q) = \neg p$$

с помощью общей таблицы значений. (Расчет более сложной операции, представленной на левой стороне тождества, становится более наглядным путем приведения в таблицу частных результатов; окончательный результат выделяется жирным шрифтом.)

		$\neg(p \wedge q) \wedge \neg(p \wedge \neg q)$					
		$\neg(p \wedge q)$			$\neg(p \wedge \neg q)$		
p	q	$p \wedge q$	$\downarrow$	$\downarrow$	$\downarrow$	$p \wedge \neg q$	$\neg p$
п	п	п	л	**л**	п	л	**л**
п	л	л	п	**л**	л	п	**л**
л	п	л	п	**п**	п	л	**п**
л	л	л	п	**п**	п	л	**п**

Ниже представлены тождества без доказательств; в их правильности читатель может убедиться с помощью построения таблицы значений.

НЕКОТОРЫЕ ДРУГИЕ ЛОГИЧЕСКИЕ ОПЕРАЦИИ

В области операций на логических переменных помимо отрицания и конъюнкции оказываются полезными некоторые другие операции.

В области одномерных логических операций фактический интерес представляет только отрицание. Ниже приведена сводная таблица значений, приводимых в следующей части операций.

p	q	(1)	(2)	(3)	(4)	(5)	(6)
п	п	п	л	п	п	л	л
п	л	п	л	л	л	п	п
л	п	п	л	п	л	п	п
л	л	л	п	п	п	л	п

ДИЗЪЮНКЦИЯ

Операция (1) называется *дизъюнкцией* и обозначается символом «$p \vee q$» (иначе ее называют альтернацией, адъюнкцией, логическим сложением), или «$p + q$». Дизъюнкция выражается с помощью операций конъюнкции и отрицания в следующей форме:

$$\boxed{p \vee q = \neg(\neg p \wedge \neg q)\,.}$$

Связь, созданная между двумя высказываниями при помощи уступительного союза «или», является такой операцией, которой в области логических значений соответствует операция дизъюнкции: высказывание является ложным тогда и только тогда, если оба высказывания ложны.

(Союз «или» в таком случае применяется в значении *допущения*, если допускается правильность обоих высказываний). Например: «выпал дождь или полили парк». Поэтому такое соединение двух высказываний также называется дизъюнкцией. (Символ «$\vee$» читается также как «или»).

Операция конъюнкция выражается с помощью операций дизъюнкции и отрицания:

$$\boxed{p \wedge q = \neg(\neg p \vee \neg q)\,.}$$

(Значение обеих сторон будет n тогда и только тогда, если $p = n = q$). Таким образом, руководствуясь теоремой, представленной на стр. 636, каждая логическая операция может быть выражена с помощью только операций дизъюнкции и отрицания.

«НИ—НИ»

Операция (2) является отрицанием операции (1) и, таким образом, выражается в виде $\neg(p \vee q)$, или в виде $\neg p \wedge \neg q$. Другое обозначение этой операции: «$p \| q$» (читается как «p два штриха q»). Данная операция иногда называется операцией «ни-ни» (из-за вида $\neg p \wedge \neg q$).

Если A и B представляют собой высказывания, то сложное высказывание типа «ни A, ни B» (которое, естественно, может быть записано в виде «Неправильно, что A и неправильно, что B») правильно тогда и только тогда, если оба высказывания A и B ложны.

Таким образом, такой операции высказывания соответствует логическая операция $p \| q$.

Отрицание также может быть выражено с помощью операции «ни-ни»:

$$\boxed{\neg p = p \| q\,,}$$

даже дизъюнкция выражается с помощью операции «ни-ни»:

$$\boxed{p \vee q = \neg\neg(p \vee q) = \neg(p \,||\, q) = (p \,||\, q) \,||\, (p \,||\, q).}$$

Таким образом, с помощью данной операции выражается каждая логическая операция (как было показано выше, каждая логическая операция описывается с помощью операций отрицания и дизъюнкции).

ИМПЛИКАЦИЯ

Операция (3) обозначается символом «$p \to q$» и называется *импликацией* (с предварительным членом p и с последующим членом q). Иное ее обозначение: $p \supset q$. Она выражается в следующем виде:

$$\boxed{p \to q = \neg(p \wedge \neg q) = \neg p \vee q.}$$

Допустим, что если $p = п$, то значение выражения $p \to q$ будет или $п$, или $л$ в зависимости от того, является ли значение q $п$, или $л$. Это аналогично тому, что высказывание типа «если A, то B», в котором первый член A является правильным, считается или правильным, или ложным в зависимости от того, правильный или ложный второй его член B. Поэтому соединению типа «если A, то B» соответствует импликация в области логических значений. Но в то же время при ложном высказывании A предложение типа «если A, то B» может вообще не считаться высказыванием. Например: если горит лампочка, то лифт работает.

Если высказывание «горит лампочка» правильно, то правильностью высказывания «лифт работает» однозначно решается правильность вышеприведенного предложения. Но если высказывание «горит лампочка» ложно, то ничего нельзя сказать о правильности вышеприведенного предложения. Можно сказать: надо подождать, пока лампочка загорится. Приведем пример, в котором не будет даже возможности «подождать»:

Если $2 \cdot 2 = 5$, то Дунай является европейской рекой. Если принять то, что соединение типа «если ... то» соответствует операции импликации, при соблюдении последнего тождества высказывание «если A, то B» выражалось бы с помощью операций конъюнкции и отрицания в следующем виде: «неправильно, что: A и не B» (здесь присутствует выражение «не B» вместо выражения «неправильно, что B»; таким образом, ясно, что выражение «неправильно, что», расположенное в начале высказывания, относится не только к A, но и к выражению «A и не B»). В соответствии с этим приведенные выше два предложения в примере могут быть переформулированы следующим образом:

а) Неправильно, что горит лампочка и лифт не работает.

б) Неправильно, что $2 \cdot 2 = 5$ и Дунай не является европейской рекой.

Если выражение «горит лампочка» ложно, то ложно и выражение «лампочка горит и лифт не работает», а отрицание его — по а) — является правильным. Выражение «$2 \cdot 2 = 5$» ложно, ложно также и выражение «Дунай не является европейской рекой»; их конъюнкция — также ложна, а отрицание этой конъюнкции — по б) — является правильным. Здесь нет противоречия по сравнению с обычным пониманием вещей, так как обычно не обращают внимание на правильность сложного предложения типа «если ... то» в том случае, когда первый член соединения является ложным.

Выражения вида «если A, то B» можно считать синонимами выражений вида «неправильно, что: «A и не B»; они называются импликациями (с предварительным членом A, с последующим членом B); для их обозначения применяется символ $A \to B$.

Представленное в области логических значений понятие импликации типа $p \to q$ соответствует понятию вышеприведенной операции высказывания.

Операции на высказываниях, выражаемые с помощью союзов и частиц, сформулированы недостаточно точно; в большинстве случаев, они до некоторой степени двусмысленны. По всей вероятности распознавание операций конъюнкции и отрицания наименее проблематично в их грамматической форме представления. Поэтому большое значение имеет возможность выражения любой логической операции через операции конъюнкции и отрицания. Как было показано выше, это позволило нам истолковать образование сложного предложения вида «если ... то» как операцию.

Упоминаются еще некоторые грамматические синонимы операции «$A \to B$»: «B, если только A», «Только тогда A, если B», «Достаточным условием B является A», «Необходимым условием A является B», «B если не A».

И конъюнкция и дизъюнкция выражаются с помощью операций импликации и отрицания:

$$\boxed{p \vee q = \neg p \to q, \quad p \wedge q = \neg(p \to \neg q).}$$

Поэтому любая логическая операция может быть выражена с помощью операций отрицания и импликации.

ЭКВИВАЛЕНТНОСТЬ

Операция (4) называется *эквивалентностью* и обозначается символом $p \leftrightarrow q$ (читается: p эквивалентно q). Выражениями данной операции являются следующие:

$$\boxed{\begin{gathered} p \leftrightarrow q = \neg(p \wedge \neg q) \wedge \neg(\neg p \wedge q) = \neg[\neg(\neg p \vee q) \vee \\ \vee \neg(p \vee \neg q)] = (p \wedge q) \vee (\neg q \wedge \neg q) = (p \to q) \wedge (q \to p). \end{gathered}}$$

Последний вид выражения операции эквивалентности объясняет ее обозначение. Другие ее обозначения: $p \equiv q$, $p \sim q$. (Конечно, последнее обозначение применяется только в тех случаях, когда отрицание не обозначается символом «$\sim$».)

Так как высказывание $p \leftrightarrow q = n$ тогда и только тогда, когда $p = q$, то данная логическая операция соответствует образованию сложного предложения вида «A тогда и только тогда, когда B». Понимание и логическое значение предложения такого характера, образованного из двух любых высказываний, иногда затруднительно для восприятия человека, как и понимание предложения вида «если ... то». Например, «$2 < 3$ тогда и только тогда, если светит солнце».

Поэтому данное предложение понимается операцией калькуляции высказываний исключительно в том случае, если считать его синонимом высказываний вида «неправильно, что A и не B, и, неправильно, что не A и B». В этом случае данная операция обозначается символом «$A \leftrightarrow B$» и называется эквивалентностью.

Часто встречаются следующие синонимы данной операции: «Для A необходимо и достаточно B», «A именно тогда, когда B».

ВЗАИМОИСКЛЮЧАЮЩЕЕ «ИЛИ»

Операция (5) является отрицанием операции (4) и, таким образом, она выражается в виде $\neg(p \leftrightarrow q)$ или $\neg(p \wedge q) \wedge \neg(\neg p \wedge \neg q)$. Для отдельного ее обозначения применяется символ $p \bigtriangledown q$.

Результатом данной операции является n тогда и только тогда, если один из членов высказывания имеет значение n. Итак, данная операция соответствует применению взаимоисключающего союза «или». Например:

Или ты вылечишься до завтрашнего дня, или мы тебя отвезем в больницу. (В соответствии с этим символ $\bigtriangledown$ читается как «или же».) В общеупотребительной терминологии часто трудно решить, в каком смысле применяется союз «или» (в допускающем или взаимоисключающем); иногда трудность состоит в том, что два высказывания, соединенные союзом «или», исключают друг друга, и поэтому нельзя принять решение, является ли окончательный вывод при правильности обоих высказываний правильным или ложным. В математической терминологии союз «или» применяется всегда в допускающем смысле, следовательно, выражается через операцию дизъюнкции. Операция $p \bigtriangledown q$ (иногда называется операцией Жегалькина) имеет значение для технического применения математической логики.

ОПЕРАЦИЯ ШЕФФЕРА

Операция (6) является отрицанием операции конъюнкции: $\neg(p \wedge q)$ или $\neg p \vee \neg q$. Отдельное ее обозначение: $p \mid q$ (читается: «p штрих q»); это операция Шеффера. Каждая логическая операция выражается с помощью этой операции (как и с помощью операции $p \parallel q$). Данная операция также имеет значение в области технических применений.

НЕКОТОРЫЕ ТОЖДЕСТВА

Выше были рассмотрены основные двухмерные операции. Представим несколько (пока не упомянутых) важных тождеств:

$p \vee p = p$	(идемпотенция дизъюнкции)
$p \vee \neg p = n$	(закон исключения третьей возможности)
$p \vee q = q \vee p$	(коммутативность дизъюнкции)
$(p \vee q) \vee r = p \vee (q \vee r)$	(ассоциативность дизъюнкции)
$p \wedge (q \vee r) = (p \wedge q) \vee (p \wedge r)$ $p \vee (q \wedge r) = (p \vee q) \wedge (p \vee r)$	(дистрибутивность)
$\neg(p \wedge q) = \neg p \vee \neg q$ $\neg(p \vee q) = \neg p \wedge \neg q$	(правила де Моргана)
$p \to q = \neg q \to \neg p$	(контрпозиция)
$(p \wedge q) \to r = p \to (q \to r)$	(экспортация)
$p \to (q \to r) = q \to (p \to r)$	
$p \wedge (p \vee q) = p$ $p \vee (p \wedge q) = p$	(законы поглощения)
$p \leftrightarrow p = n$	
$\neg(p \leftrightarrow q) = \neg p \leftrightarrow q = p \leftrightarrow \neg q$	
$p \leftrightarrow q = q \leftrightarrow p$	

Свойства коммутативности и ассоциативности дизъюнкции дают возможность записывать операцию повторной дизъюнкции без скобок как «многочленную дизъюнкцию».

В ходе доказательства обычно встречаются следующие операции калькуляции высказываний: отрицание, конъюнкция, дизъюнкция, импликация и эквивалентность. Больше двух операций почти не бывает за исключением операций многочленной конъюнкции и многочленной дизъюнкции. Поэтому можно опустить анализ логических операций с числом переменных, превышающим два. Для таких операций не введены отдельные обозначения и знаки операций.

ПОНЯТИЕ СЛЕДСТВИЯ В КАЛЬКУЛЯЦИИ ВЫСКАЗЫВАНИЙ

ФОРМУЛЫ КАЛЬКУЛЯЦИИ ВЫСКАЗЫВАНИЙ

До сих пор рассматривались соотношения *алгебраического* характера логических операций; знание их позволяет объяснить понятие правильного вывода в области калькуляции высказываний. Сначала введены некоторые названия:

Формулами калькуляции высказываний называются следующие:

а) логические значения *п* и *л* и логические переменные;

б) последовательность знаков, полученная из (предварительно заданной) формулы путем постоянного применения знака $\neg$; т. е. выражение $\neg(\alpha)$ называется формулой, если выражение α является формулой;

в) последовательность знаков, возникшая из двух (предварительно заданных) формул путем постоянного применения одного из знаков $\wedge, \vee, \rightarrow, \leftrightarrow$; т. е. формулы $(\alpha) \wedge (\beta)$, $(\alpha) \vee (\beta)$, $(\alpha) \rightarrow (\beta)$, $(\alpha) \leftrightarrow (\beta)$ называются формулами, если α и β являются формулами.

Скобки обычно опускаются: например, вместо выражения $(p) \vee (q)$ пишется $p \vee q$.

Другие последовательности знаков не считаются формулами калькуляции выскаываний.

Некоторая формула называется тавтологией (тождественно правильной, действительной), если в ее таблице значений (в столбце результата) присутствуют только одни знаки *п*.

Такое свойство формулы α, что она является тавтологией, обозначается символом «$\vdash \alpha$». Представим некоторые примеры тавтологии:

$$\vdash п; \quad \vdash p \vee \neg p; \quad \vdash (p \wedge q) \leftrightarrow (p \wedge p).$$

Формула называется удовлетворимой, если ее таблица значений содержит хотя бы один знак *п*.

Примеры удовлетворимых формул:

$$p \wedge q, \quad p \rightarrow q, \quad p \vee q, \; p \rightarrow q, \; \neg p.$$

Формула называется контрадикцией (тождественно ложной, неудовлетворимой), если таблица ее значений содержит одни знаки *л*.

Примеры:

$$л, \quad p \wedge \neg p, \quad p \leftrightarrow \neg p; \quad \neg(p \vee \neg p).$$

Формула является тавтологией тогда и только тогда, если ее отрицание — контрадикция.

Две формулы α и β называются равноценными, если по таблице значений они выражают тождественные операции, (т. е. в соответствии с правилами, приведенными выше, если $\alpha = \beta$). По таблице значений операции эквивалентности выражение $\alpha \leftrightarrow \beta$ представляет собой тавтологию. Итак, равноценность α и β выражается символом «$\vdash \alpha \leftrightarrow \beta$».

ЗАМЕНА И ДОПОЛНЕНИЕ

Если вместо некоторой переменной формулы α подставляется любая, но зафиксированная формула β, то о полученной таким образом формуле говорится, что она возникла из формулы α путем замены.

Пример: Заменим переменную p формулы $\neg p \to (p \to q)$ формулой $p \vee q$. Результат будет следующий:

$$(1) \qquad \neg(p \vee q) \to ((p \vee q) \to q).$$

Как исходная формула, так и результат представляет собой тавтологию.

Путем замены из *тавтологии* (или из *контрадикции*) получится *тавтология* (или *контрадикция* соответственно). (Удовлетворимая формула после применения к ней замены не обязательно остается удовлетворимой. Например, если в удовлетворимой формуле $p \wedge q$ переменная p заменяется формулой $q \wedge \neg q$, то в результате получится контрадикция $q \wedge \neg q \wedge q$.) По определению, формула β является *частью* формулы α, если формула α имеет вид «. . . β —», где на место знаков «. . .» и «—» могут быть поставлены и другие знаки (допуская при этом, что ничего не стоит на первом, или на втором, или на обоих местах). Например, частью формулы $(p \wedge q) \to (q \leftrightarrow \neg r)$ является сама формула, ее следующие (действительные) части: $(p \wedge q)$, $(q \leftrightarrow \neg r)$, $\neg r$ и т. д.; но выражение «$p \wedge \neg r$» не является ее частью, так как, хотя каждый член этого выражения присутствует в этой формуле, но не связно. Выражение «$\to (q \leftrightarrow$» также не является частью оригинальной формулы, так как, хотя оно и присутствует в связном виде в оригинальной формуле, но не представляет собой формулу.

Если подставить формулу γ вместо некоторой части β формулы α, то о полученной таким образом формуле говорится, что она получена из формулы α путем дополнения. Дополнение считается равноценным, если $\vdash \beta \leftrightarrow \gamma$, т. е. если часть формулы дополняется равноценной ей частью.

Например: дополнив часть $\neg(p \vee q)$ в формуле (1) равноценной ей формулой $\neg p \wedge \neg q$, получим следующую формулу:

$$(2) \qquad (\neg p \wedge q) \to ((p \vee q) \to q).$$

Как формула (1) так и (2) являются тавтологиями.

Формула, полученная путем применения равноценного дополнения, является равноценной оригинальной формуле. Таким образом, путем применения равноценного дополнения из тавтологии (или из контрадикции, или из удовлетворимой формулы) получится тавтология (или контрадикция или удовлетворимая формула соответственно).

СХЕМЫ СЛЕДСТВИЯ ПО КАЛЬКУЛЯЦИИ ВЫСКАЗЫВАНИЙ

По определению, следствием формул $\alpha_1, \alpha_2, \ldots, \alpha_n$ является формула β (в смысле калькуляции высказываний), если в общей таблице значений этих формул нет такой строки, в которой значениями каждой из формул $\alpha_1, \ldots, \alpha_n$ являются знаки n, а значением формулы β — $л$.

(Во всех случаях, если каждая из переменных $\alpha_1, \ldots, \alpha_n$ «правильна», и формула β — «привильна». Но если все переменные $\alpha_1, \ldots, \alpha_n$ не могут быть «правильными» одновременно, формула β и в таком случае считается их следствием.)

Такое соотношение обозначается символом «$\alpha_1, \ldots, \alpha_n \vdash \beta$»; данный символ называется «схемой следствия», формулы $\alpha_1, \ldots, \alpha_n$ — предпосылками данной схемы, формула β — окончательным выводом схемы.

Более традиционный вид такой схемы:

$$\frac{\alpha_1, \ldots, \alpha_n}{\beta}.$$

Применяется также и символ «$\alpha_1, \ldots, \alpha_n \to \beta$» (при условии, если операция импликации не обозначается символом $\to$).

Поскольку все формулы $\alpha_1, \ldots, \alpha_n$ могут иметь значение n только в том случае, если выполняется условие $\alpha_1 \wedge \ldots \wedge \alpha_n = n$, то обе схемы следствия «$\alpha_1, \ldots, \alpha_n \vdash \beta$» и «$\alpha_1 \wedge \ldots \wedge \alpha_n \vdash \beta$» являются правильными одновременно. Основываясь на вышеприведенном рассуждении, достаточно заниматься схемами следствия «с одной предпосылкой».

Поскольку формула $\alpha \to \beta$ является тавтологией тогда и только тогда, когда для всех оценок присутствующих в ней переменных, при условии $\alpha = n$ одновременно выполняется и условие $\beta = n$, то *одновременно правильны* соотношения «$\alpha \vdash \beta$» и «$\vdash \alpha \to \beta$». Следовательно, можно сформулировать результат: *Следствием формулы α является формула β* (в смысле понятий по калькуляции высказываний), *если выражение $\alpha \to \beta$ является тавтологией.*

Ниже представлены некоторые схемы следствия:

$p;\ p \to q \vdash q$	(отделение; по-латински: modus ponens)
$p \to q;\ \neg q \vdash \neg p$	(косвенное опровержение)
$\neg p \to q;\ \neg q \vdash p$	(косвенное доказательство)
$p \to q;\ p \to \neg q \vdash \neg p$	(сокращение до абсурда)
$p \vee q;\ \neg p \vdash q$	(modus tollens)
$p \to q;\ q \to r \vdash p \to r$	(цепное правило)

Применение замены и равноценного дополнения приведет к (правильной) схеме следствия из (правильной) схемы следствия (конечно, операцию замены следует проводить во всех предпосылках и в конечном выводе).

Из схем следствия чаще всего применяется *отделение* (мы еще вернемся к проблеме значения этой схемы следствия).

Знаки *п* и *л* могут быть устранены из формул (например, могут быть заменены выражениями $p \vee \neg p$ или $p \wedge \neg p$).

Под интерпретацией некоторой формулы α подразумевается такое высказывание A, которое получается из формулы α путем замены (или, может быть, преобразования) присутствующих в ней переменных на высказывания (но сначала устраняются знаки *п* и *л*, если они присутствовали в исходной формуле).

Путем интерпретации предпосылок и окончательного вывода некоторой (правильной) схемы следствия получится (правильный по понятиям калькуляции высказываний) вывод.

По такому принципу в области калькуляции высказываний вывод, считающийся правильным, контролируется следующим образом. Все высказывания, присутствующие в данном выводе, переформулируются так, чтобы все, представленные в выводе операции и их члены, стали однозначно распознаваемыми (также переформулируются все операции, отличающиеся от операций $\neg$, $\wedge$, $\vee$, $\rightarrow$, $\leftrightarrow$; если в выводе встречаются такие операции), записываются знаки распознанных операций и скобки, указывающие границы их действия. Затем применяется такой же метод для членов операций, и процесс продолжается до тех пор, пока не получатся так называемые элементарные высказывания, в которых уже нет операций калькуляции высказываний. Наконец, элементарные высказывания заменяются логическими переменными (одинаковые — одинаковыми, различные — различными переменными). В конечном итоге вместо каждого высказывания получится по одной формуле; естественно, исходные высказывания являются интерпретациями этих формул. Исследуемому выводу соответствует схема следствия, правильность которой контролируется, например, с помощью таблицы значений.

ПРИМЕР ПРИМЕНЕНИЯ

Применение вышеприведенного метода показывается на одном примере. Пусть контролируется правильность следующего «текстового» вывода:

Если транспорт задерживается, или если погода неблагоприятная, то не будет ни выполнения плана, ни раздачи премии и участия в прибылях, если только бригада Вёрёш не займет твердую позицию и возьмется за дело.

Если погода неблагоприятная и если бригада Вёрёш не займет твердую позицию, то не будет выполнения плана.

Преобразование предпосылки имеет следующий вид: (Транспорт задерживается $\vee$ погода неблагоприятная) $\rightarrow$ {[(будет выполнение плана) $\wedge \neg$ (будет раздача премии $\wedge$ будет участие в прибылях)] $\vee$ [бригада Вёрёш займет твердую позицию $\wedge$ бригада Вёрёш возьмется за дело]}.

Преобразование окончательного вывода:

[Погода неблагоприятная $\wedge \neg$ (бригада Вёрёш займет твердую позицию)] $\rightarrow \neg$ (будет выполнение плана).

Элементарные высказывания заменяются логическими переменными:

p: «транспорт задерживается»
q: «погода неблагоприятна»
r: «будет выполнение плана»
s: «будет раздача премии»
t: «будет участие в прибылях»
u: «бригада Вёрёш возьмется за дело».
v: «бригада Вёрёш займет твердую позицию»

Итак, предпосылкой является интерпретация следующей формулы:

$$(p \vee q) \to \{[\neg r \wedge \neg (s \wedge t)] \vee [u \wedge v]\},$$

а окончательным выводом —

$$(q \wedge \neg u) \to \neg r .$$

Следовательно, нужно контролировать правильность следующей схемы:

$$(p \vee q) \to \{[\neg r \wedge \neg (s \wedge t] \vee [u \wedge v]\} \vdash (q \wedge \neg u) \to \neg r .$$

Вместо составления таблицы значений применим метод косвенного рассуждения. Предположим, что настоящая схема неправильна, т. е. имеется такая оценка переменных, при которой предпосылка — n, а окончательный вывод — $л$. Окончательный вывод как импликация может иметь значение $л$, если $q \wedge \neg u$ n и $\neg r = л$, а такая ситуация возможна только при значениях $q = n$, $u = л$, $r = n$. Если подставить эти значения в предпосылку, получится следующее соотношение:

$$(p \vee n) \to \{[л \wedge \neg (s \wedge t)] \vee [л \wedge v]\} .$$

Но $p \vee n = n$, $л \wedge \neg (s \wedge t) = л$, $л \wedge v = л$, и, таким образом, из предпосылки получится следующее:

$$n \to (л \vee л) = n \to л = л ;$$

отсюда: при значении окончательного вывода $л$, значение предпосылки также будет $л$. Итак, исключается такой случай, при котором предпосылка — n, а окончательный вывод — $л$: данная схема следствия правильна. Нетрудно определить, что если $\vdash \alpha \leftrightarrow \beta$, то $\alpha \vdash \beta$ и $\beta \vdash \alpha$. На основе этого введен символ $\alpha \dashv\vdash \beta$ для обозначения равноценности формул α и β.

Ясно, что схема следствия, представленная в вводной части настоящей главы, является, по сути дела, схемой отделения.

Но понятие следствия по калькуляции высказываний оказывается недостаточно общим даже при анализе некоторых простых выводов, считающихся правильными по «здравому смыслу»:

Каждая дифференцируемая функция является непрерывной.
Каждая непрерывная функция является интегрируемой.

Каждая дифференцируемая функция является интегрируемой.

Эти высказывания представляют собой элементарные высказывания (они не делятся дальше). Следовательно, данному выводу соответствовала бы схема типа «$p, q \vdash r$», но она, конечно, является неправильной схемой следствия. Вышеприведенный и аналогичные выводы рассматриваются в области *калькуляции предикатов.*

«АКСИОМАТИЧЕСКОЕ» ПОСТРОЕНИЕ КАЛЬКУЛЯЦИИ ВЫСКАЗЫВАНИЙ

Может быть доказано, что любая тавтология может быть воспроизведена из следующих тавтологий путем применения конечного числа операций *замены и отделения:*

1. $p \to (q \to p)$
2. $(p \to q) \to \{[p \to (q \to r)] \to [p \to r]\}$
3. $p \to [q \to (p \wedge q)]$
4. $(p \wedge q) \to p$
5. $(p \wedge q) \to q$
6. $p \to (p \vee q)$
7. $q \to (p \vee q)$
8. $(p \to r) \to](q \to r) \to ((p \vee q) \to r)]$
9. $(p \to q) \to [(p \to \neg q) \to \neg p]$
10. $\neg \neg p \to p$

Понятие отделения подразумевает следующее: если формула вида $\alpha \to \beta$ и формула α уже воспроизведены, то в соответствии со схемой «$\alpha;\ \alpha \to \alpha \vdash \beta$» формула β также воспроизведена.

Пример: Пусть воспроизводится тавтология $(p \wedge q) \to (p \wedge q)$. Во второй аксиоме вместо q подставляется $q \to p$ и вместо r — p:

$$[p \to (q \to p)] \to \{[p \to ((q \to p) \to p)] \to (p \to p)\}\,.$$

Предварительным членом настоящего соотношения является именно первая аксиома. Эта часть может быть отделена.

(a) $$[p \to ((q \to p) \to p)] \to (p \to p)\,.$$

Заменим q $q \to p$ в 1-ой аксиоме:

$$p \to ((q \to p) \to p)\,.$$

Данное соотношение является предварительным членом соотношения (а), следовательно, может быть отделено. Остается следующее выражение:

$$p \to p\,.$$

Если в данном соотношении $p \wedge q$ заменить p, получится искомая формула.

Поскольку путем применения операций замены и отделения тавтология получается из тавтологии, то приведенный выше метод так же хорошо может быть использован для распознавания тавтологий (и для обоснования понятия вывода), как калькуляция логических значений, метод, основанный на применении таблицы значений. Представленный выше метод называется «аксиоматическим построением» калькуляции высказываний. Тавтологии 1—10 представляют собой «аксиомы», а замена и отделение — «правила выведения». В области калькуляции высказываний известно несколько других подходов «аксиоматического построения», с другими «аксиомами», «правилами выведений». В то же время понятие аксиоматического построения имеет значение только при калькуляции предикатов.

Применение вышеприведенного аксиоматического построения показывает особое значение правила отделения, так как с помощью его (и замены) каждая тавтология и, следовательно, каждая схема следствия (в виде тавтологии $\alpha \to \beta$) могут быть выведены из некоторых заданных тавтологий.

В выщеприведенной системе калькуляции высказываний операция эквивалентности не присутствует; здесь формулы вида $\alpha \leftrightarrow \beta$ могут применяться как сокращения формул вида $(\alpha \to \beta) \wedge (\beta \to \alpha)$.

КАЛЬКУЛЯЦИЯ ПРЕДИКАТОВ

ПРЕДИКАТЫ И КВАНТОРЫ

При анализе вывода, приведенного в примере в предыдущей главе (и при анализе многих других выводов подобного типа) следует отметить, что применяемые нами высказывания могут быть приведены из так называемых *открытых предложений* или *предикатов*, примерами которых служат следующие:

... является неделимым числом;
... является столицей ...;
... купил ... за ... рублей.

Если в эти схемы предложений вставить названия соответственно подобранных предметов (вместо пунктира), то получатся *замкнутые предложения, высказывания*. Такие предикаты в некоторых случаях выражаются однозначно, если вместо пунктира записываются буквы (x, y, z и т. д.). В приведенном примере путем применения буквенного обозначения вместо пунктира получится следующая форма:

x является неделимым числом;
y является столицей x;
x купил y за z рублей.

С помощью такой формы обозначений вместо пунктира можно записать одинаковые или различные названия. Следующие два предиката имеют разные значения:

x попросил y, чтобы он помогал z;
x попросил y, чтобы он помогал x.

Но если применить пунктирную форму обозначения, эти два предиката нельзя различить:

... *попросил* ..., *чтобы он помогал* ...

Кроме заполнения оставленных свободными (пунктиром или буквенным обозначением) мест названиями имеется и другой способ образования высказываний из предикатов: *квантификация*. Например, из открытого предложения «если x представляет собой дифференцируемую функцию, то функция x — непрерывная функция», подставив перед предложением «Для каждого x», получим следующее: Для каждого x: если x представляет собой дифференцируемую функцию, то x представляет собой непрерывную функцию.

В устной речи это выражается так:

«Каждая дифференцируемая функция является непрерывной.»

Закрытие открытого предложения таким способом называется «*универсальной квантификацией*». Текст «Для каждого x» обозначается символом «$\forall x$» (вместо x может стоять и другая буква) и называется универсальным квантором.

Если в открытом предложении представлено несколько переменных (с обозначениями x, y, z и т. д.), то из него получается высказывание только путем многократного применения универсальной квантификации (для каждой переменной приходится повторять процесс квантификации).

Из открытого предложения «x является непрерывной функцией и x нигде не дифференцируема» путем поставления перед ним текста «Имеется такое x, что», получится следующее предложение:

Имеется такое x, что x является непрерывной функцией и x нигде не дифференцируема.

В устной речи это выражается так: «Имеется такая непрерывная функция, которая нигде не дифференцируема» или (по формулировке, являющейся традиционно несколько обманчивой в области классической логики: «Некоторые непрерывные функции нигде не являются дифференцируемыми». Такой вид закрытия открытого предложения носит название *экзистенциальная квантификация*.

Текст «Имеется такое x» обозначается символом «$\exists x$» и называется экзистенциальным квантором.

Для точного анализа вводятся следующие понятия.

Под понятием n-мерного предиката ($n \geq 1$), определенного на непустом множестве H, подразумевается такая функция, областью существования которой является множество упорядоченных n-членных знаков, образуемых из элементов множества H, а значениями — логические значения.

Логические значения считаются 0-мерными предикатами. Предикаты обозначаются символами $\mathbf{F}x$, $\mathbf{G}xy$, $\mathbf{H}x_1 x_2 \ldots x_n$ и т. д.; жирными буквами обозначаются предикаты, а строчными буквами — аргументы предиката как функции; количеством последних определяется размерность предиката (сколько в нем переменных).

Примеры: Пусть H представляет собой множество натуральных чисел и предикат $\mathbf{F}x$ определяется таким образом:

$$\mathbf{F}x = \begin{cases} п, & \text{если } x \text{ является неделимым числом,} \\ л & \text{в любом другом случае.} \end{cases}$$

Это определение предиката «неделимого числа».

Пусть H представляет собой множество стран и столиц Земли, а предикат $\mathbf{G}xy$ — следующее:

$$\mathbf{G}xy = \begin{cases} п, & \text{если } (y \text{ — столица, } x \text{ — страна}) \; y \text{ является столицей страны } x, \\ л & \text{в любом другом случае.} \end{cases}$$

Это предикат термина «столица».

Из некоторого (по меньшей мере двухмерного) предиката образуется предикат, число переменных которого на 1 меньше, чем у исходного предиката, если предписать тождество двух членов, расположенных на зафиксированных местах упорядоченных n-членных знаков, являющихся возможными аргументами. Например, из трехмерного предиката $\mathbf{F}xyz$ может быть образован двухмерный предикат $\mathbf{F}xyx$.

Частный пример: Пусть H представляет собой множество людей и

$$\mathbf{F}xyz = \begin{cases} п, & \text{если } x \text{ попросил } y, \text{ чтобы он помогал } z, \\ л & \text{в любом другом случае} \end{cases}$$

Тогда:

$$\mathbf{F}xyx = \begin{cases} п, & \text{если } x \text{ попросил помощи (для себя) у } y, \\ л, & \text{в любом другом случае.} \end{cases}$$

ОПЕРАЦИИ С ПРЕДИКАТАМИ

Для предикатов, определенных на одном и том же множестве, может быть представлено и понятие логических операций.

Например: под термином конъюнкции предикатов $\mathbf{F}x_1 \ldots x_k$ и $\mathbf{G}y_1 \ldots y_n$ подразумевается такой $(k+n)$-мерный предикат, значением которого является n для упорядоченного $(k+n)$-членного знака $(x_1, \ldots, x_k, y_1, \ldots, y_n)$ тогда и только тогда, если $\mathbf{F}x_1 \ldots x_k = n$ и $\mathbf{G}y_1 \ldots y_n = n$; это обозначается символом $\mathbf{F}x_1 \ldots x_k \wedge \mathbf{G}y_1 \ldots y_n$.

Если же имеются тождественные среди переменных $x_1, \ldots, x_k, y_1, \ldots, y_n$, то предикат, полученный таким образом, имеет переменных меньше чем $(k+n)$. Например, предикат $\mathbf{F}x \wedge \mathbf{G}yz$ является трехмерным, предикат $\mathbf{F}x \wedge \mathbf{G}yx$ — двухмерным, а предикат $\mathbf{F}x \wedge \mathbf{G}xx$ — одномерным. Аналогично объясняются и другие логические операции.

Частные примеры: Пусть H представляет собой некоторое множество людей и рассматриваются следующие предикаты:

$$\mathbf{F}x = \begin{cases} n, & \text{если } x \text{ является больным,} \\ л & \text{в любом другом случае.} \end{cases}$$

$$\mathbf{G}xy = \begin{cases} n, & \text{если } x \text{ лечит } y, \\ л & \text{в любом другом случае.} \end{cases}$$

Тогда:

$$\mathbf{F}x \wedge \mathbf{G}xx = \begin{cases} n, & \text{если } x \text{ болен и лечит себя,} \\ л & \text{в любом другом случае;} \end{cases}$$

$$\mathbf{F}y \wedge \mathbf{G}xy = \begin{cases} л, & \text{если } y \text{ болен и } x \text{ его не лечит,} \\ n & \text{в любом другом случае.} \end{cases}$$

Если $\mathbf{F}xy$ представляет собой двухмерный предикат, символом $\forall x\, \mathbf{F}xy$ обозначается такой одномерный предикат, значение которого тогда и только тогда является n для некоторой переменной a множества H, если при любой паре (x, a) выполняется условие $\mathbf{F}xa = n$. Аналогично истолковывается применение универсального квантора для любой переменной любого, по меньшей мере двухмерного предиката.

Данной операцией воспроизводится предикат, число переменных которого на одно меньше, чем число переменных исходного предиката.

Пример: Если $\mathbf{G}xy$ представляет собой предикат «лечит», приведенный в последнем примере, тогда действительно следующее соотношение:

$$\forall y\, \mathbf{G}xy = \begin{cases} n, & \text{если } x \text{ лечит каждый элемент множества } H, \\ л & \text{в любом другом случае.} \end{cases}$$

Наконец, применение квантора $\forall x$ для одномерного предиката объясняется таким образом, что результатом данной операции является 0-мерный предикат, т. е. некоторое логическое значение: $\forall x\, \mathbf{F}x = n$ тогда и только тогда, если $\mathbf{F}x = n$ для любого элемента x множества H.

Например, если $\mathbf{F}x$ представляет собой вышеприведенный предикат «болен», тогда

$$\forall x \mathbf{F}x = \begin{cases} n, & \text{если каждый элемент множества } H \text{ болен,} \\ л & \text{в любом другом случае.} \end{cases}$$

Дуальное объяснение дается для применения экзистенциального квантора: $\exists x\, \mathbf{F}xa = л$ тогда и только тогда, если при любой паре (x, a) выполняются условия $\mathbf{F}xa = л$ и $\exists x\, \mathbf{F}x = л$, если при любом значении x выполняется условие $\mathbf{F}x = л$.

Следовательно, если $\mathbf{G}xy$ представляет, например, предикат «лечит», приведенный уже несколько раз, то:

$$\exists x \mathbf{G}xy = \begin{cases} n, & \text{если, по меньшей мере, один из элементов} \\ & \text{множества } H \text{ лечит } y \\ л & \text{в любом другом случае;} \end{cases}$$

и

$$\exists x\, \forall y \mathbf{G}xy = \begin{cases} n, & \text{если имеется хотя бы один такой элемент в множестве} \\ & H, \text{ который лечит каждый элемент множества } H, \\ л & \text{в любом другом случае;} \end{cases}$$

но с другой стороны:

$$\forall x\, \exists x \mathbf{G}xy = \begin{cases} n, & \text{если хотя бы один элемент лечит любой элемент} \\ & \text{множества } H, \\ л & \text{в любом другом случае.} \end{cases}$$

В последних двух примерах наблюдается важное влияние замены порядка применения универсального и экзистенциального кванторов, действующее в сторону существенного изменения смысла.

Применение квантора $\forall x$ (или $\exists x$) для предиката, не зависящего от значения x (включая и логические значения), рассматривается как операция, оставляющая предикат неизменным. Например, $\forall x\, \mathbf{F}y = \mathbf{F}y$ и $\exists xn = n$.

Для обозначения универсального квантора применяются символы «(x)», «$\bigwedge x$», «Πx» вместо символа $\forall x$, а для обозначения экзистенциального квантора — символы «(Ex)», «$\bigvee x$», «Σx» вместо символа $\exists x$. Вместо названия «предикат» иногда применяют название «логическая функция».

ФОРМУЛЫ КАЛЬКУЛЯЦИИ ПРЕДИКАТОВ

В области калькуляции высказываний при анализе оказалось достаточным вместо самих высказываний заниматься логическими переменными. В приведенном ниже рассуждении для изображения конструкции предикатов и возникших из них (путем применения квантора) высказываний предикаты и их переменные (аргументы) заменяются соответствующими знаками. Для этого введено понятие *переменных предикатов* и *переменных индивидуума*. Переменные индивидуума обозначаются буквами x, y, z, x_1, x_2 и т. д. и (если это не вызывает недоразумения) называются просто *переменными*. Переменные предикатов обозначаются буквами F, G, H, F_1, F_2 и т. д. Для каждой переменной предиката предписывается положительное целое число, названное *числом аргументов* переменных предиката. (Во избежание недоразумений две переменные предикатов с различными числами аргументов не обозначаются одинаковыми буквами в одном и том же соотношении.) Если следует указать число аргументов, то оно ставится в верхнем индексе. Например, $F^{(2)}$ обозначает переменную предиката с двумя аргументами.

Приведенная в настоящей главе теория имеет название *калькуляция предикатов*. (Другие названия: *калькуляция логических функций*, *теория квантификации.)*

Интерпретация формул калькуляции предикатов:

а) Логические значения и логические переменные являются формулами.

б) Если F представляет собой переменную предикатов с n аргументами и $x_1, \ldots, x_n$ — переменные индивидуума, то $Fx_1 \ldots x_n$ является формулой.

в) Если α — формула, то и $\neg(\alpha)$ — формула.

г) Если α и β являются формулами, то и $(\alpha) \wedge (\beta)$, $(\alpha) \vee (\beta)$, $(\alpha) \to (\beta)$ и $(\alpha) \leftrightarrow (\beta)$ являются формулами.

д) Если α представляет собой формулу, и x — переменную предиката то и $\forall x(\alpha)$, и $\exists x(\alpha)$ представляют собой формулы.

Под термином *область действия* квантора $\forall$ и $\exists$ в формуле вида $\forall x(\alpha)$ или $\exists x(\alpha)$ подразумевается формула α. Если α является формулой вида $Fx_1 \ldots x_n$ или $\forall y(\alpha)$ и $\exists y(\beta)$, то ее не следует заключать в скобки при использовании квантора. Например, в формуле $\forall x\, \forall y\, Fxy$ областью действия $\forall y$ является Fxy, а областью действия $\forall x$ — формула $\forall y\, Fxy$; в формуле $\exists y(\forall x\, Fx \to Gyz)$ областью действия формулы $\forall x$ является Fx, а областью действия $\exists y$ — формула $\forall x Fx \to Gyz$.

В некоторой формуле одна и та же переменная может встречаться несколько раз.

Среди возможностей появления некоторой переменной необходимым считается появление переменной в кванторе и присутствие данной переменной в области действия квантора, содержащего такую же переменную.

Например, присутствие формулы x считается необходимым в формулах $\forall x$ и $\exists x$ и, кроме того, — в области действия этих же кванторов.

Другие (необязательные) возможности появления некоторой переменной называются свободным появлением переменной.

В формуле $\forall x Fx \to Gyz$ присутствие x является необходимым в обоих случаях наступления; наступление y и z — свободно; в формуле

$\exists y(\forall x\, Fx \to Gyz)$ присутствие x и y везде необходимо, а присутствие z — свободно; в формуле $\forall x\, Fx \to Gxy$ присутствие x в предварительном члене является необходимым, а в последующем члене и x, и y присутствуют свободно.

Формула называется свободной, если в ней присутствует свободно наступающая переменная. В другом случае формула называется закрытой. Если в некоторой формуле α всеми (и) свободно наступающими переменными являются $x_1, \ldots, x_n$, то под понятием универсального закрытия формулы α подразумевается формула «$\forall x_1, \ldots, \forall x_n(\alpha)$; (это, конечно, закрытая формула; порядок кванторов $\forall x_1, \ldots, \forall x_n$ не имеет значения.)

ПОНЯТИЕ СЛЕДСТВИЯ КАЛЬКУЛЯЦИИ ПРЕДИКАТОВ

Под определением интерпретации переменной предиката $F^{(n)}$ на непустом множестве H подразумевается n-мерный предикат, определенный на множестве H.

Если множество H имеет больше, чем один элемент, то переменная $F^{(n)}$ интерпретируется по-разному.

Под термином оценки некоторой закрытой формулы подразумевается следующее. Задается некоторое непустое множество H, на нем интерпретируются все переменные предиката, присутствующие в множестве H. Затем переменные предиката заменяются их интерпретациями, а если имеются логические переменные, они заменяются логическими значениями (не меняя порядок и обозначение переменных индивидуума). Полученным таким образом выражением воспроизводится логическое значение, исчислимое последовательно на основе определений предикатов и операций в выражении. Полученное значение (*п* или *л*) считается результатом данной оценки.

Пример: Пусть оценивается формула $\forall y[\exists x(Fx \wedge Gyx) \to \exists x(Hx \wedge Gyx)]$. (Данная формула представляет собой формализованный вид вывода в примере следствия на стр. 659). Пусть будет $H = \{1, 2\}$, а интерпретацией выражений F, G и H являются предикаты, приведенные ниже в таблице, $\mathbf{F}x$, $\mathbf{G}yx$ и $\mathbf{H}x$. Следует вычислить значение выражения

$$\forall y[\exists x(\mathbf{F}x\ \mathbf{G}yx) \to \exists x(\mathbf{H}x \wedge \mathbf{G}yx)]\,.$$

					$\exists x(\mathbf{F}x \wedge \mathbf{G}yx) \to \exists x(\mathbf{H}x \wedge \mathbf{G}yx)$				
					$\exists x(\mathbf{F}x \wedge \mathbf{G}yx$			$\exists x(\mathbf{H}x \wedge \mathbf{G}yx)$	
y	x	$\mathbf{F}x$	$\mathbf{G}yx$	$\mathbf{H}x$	$\mathbf{F}x \wedge \mathbf{G}yx$	$\downarrow$	$\downarrow$	$\downarrow$	$\mathbf{H}x \wedge \mathbf{G}yx$
1	1	*п*	*п*	*л*	*п*	*п*	**п**	*п*	*л*
1	2	*л*	*п*	*п*	*л*				*п*
2	1	*п*	*п*	*л*	*п*	*п*	**л**	*л*	*л*
2	2	*л*	*л*	*п*	*л*				*л*

В таблице подробно представлен способ расчета. Окончательный результат указан жирным шрифтом в соответствующем столбце. Очевидно, что значением выражения $\exists x(\mathbf{F}x \wedge \mathbf{G}yx) \to \exists x(\mathbf{H}x \wedge \mathbf{G}yx)$ является *п* при $y = 1$, и — *л* при $y = 2$. По определению универсального квантора может быть получена оценка $\forall y[\exists x(\mathbf{F}x \wedge \mathbf{G}yx) \to \exists x(\mathbf{H}x \wedge \mathbf{G}yx)] =$ *л*. Следовательно, результат такой оценки данной формулы — *л*.

Закрытая формула называется тавтологией, если результат каждой ее оценки — *п*. Открытая формула тогда называется тавтологией, если ее универсальное закрытие является тавтологией. Считается, что формула β является следствием формулы α, если соотношение $\alpha \to \beta$ представляет собой тавтологию.

Тавтология и следствие обозначаются точно так же, как и в калькуляции высказываний.

Например: $\forall x\, Fx \vdash \exists y\, Fy$, т. е. $\vdash \forall x\, Fx \to \exists y\, Fy$. Ведь если F интерпретируется предикатом $\mathbf{F}x$, имеющим тождественно значение *п* на некотором (непустом) множестве H, то $\forall x\, \mathbf{F}x =$ *п* и $\exists y\, \mathbf{F}y =$ *п* (так как множество H — непустое) и, таким образом, $\forall x\, \mathbf{F}x \to \exists y\, \mathbf{F}y =$ *п* $\to$ *п* $=$ *п*. А если F интерпретируется предикатом $\mathbf{F}x$, не имеющим тождественно значение *п*, то $\forall x\, \mathbf{F}x =$ *л*, и поэтому $\forall x\, \mathbf{F}x \to \exists y\, \mathbf{F}y =$ *п* (так как импликация с *л* предварительным членом всегда имеет значение *п*).

Другой пример: если α представляет собой формулу, в которой присутствуют по крайней мере переменные x и y свободно, то имеет место соотношение $\forall x\, \forall y(\alpha) \dashv\vdash \forall y\, \forall x(\alpha)$, т. е. $\vdash (\forall x\, \forall y(\alpha)) \to (\forall y\, \forall x(\alpha))$. Ибо если переменные предикатов в данной формуле интерпретируются путем замены таким предикатом $\mathbf{F}xy$, в отношении которого действительно соотношение $\forall x\, \forall y\, \mathbf{F}xy =$ *п* при любом значении x, то $\forall y\, \mathbf{F}xy =$ *п* будет при любом значении x и $\mathbf{F}xy =$ *п* для любой пары (x, y). Но тогда $\forall x\, \mathbf{F}xy =$ *п* при каждом значении y и $\forall y\, \forall x\, \mathbf{F}xy =$ *п*, следовательно, $(\forall x\, \forall y\, \mathbf{F}xy) \leftrightarrow (\forall y\, \forall x\, \mathbf{F}xy) =$ *п* $\leftrightarrow$ *п* $=$ *п*. Таким же образом проводится обратное рассуждение.

С помощью аналогичных рассуждений можно убедиться в том, что из каждой тавтологии калькуляции высказываний получится тавтология, если заменить логические переменные формулой калькуляции предикатов.

Также можно убедиться в правильности следующих утверждений:

$$\boxed{\vdash \forall x\, Fx \to Fy; \qquad \vdash Fy \to \exists x\, Fx; \qquad \neg \forall x\, Fx \dashv\vdash \exists x \neg Fx.}$$

ОПЕРАЦИЯ ЗАМЕНЫ В КАЛЬКУЛЯЦИИ ПРЕДИКАТОВ

Возможны следующие варианты *замены* в калькуляции предикатов: (а) замена переменной индивидуума другой переменной индивидуума; (б) замена логической переменной или переменной предиката любой формулой. Путем замены в области калькуляции предикатов воспроизводится тавтология из тавтологии (т. е. правильный вывод из правильного вывода) при соблюдении определенных правил. Эти правила здесь не приводятся подробно, суть их заключается в следующем: запрещается каждая такая замена, в результате которой некоторое свободное наступление переменной индивидуума становится обязательным (либо в формуле, в которой проводится замена, либо в формуле, с помощью которой проводится замена). Так, например, в формуле $\forall x Fxy$ запрещается замена x переменной y, так как в формуле $\forall y\, Fxy$ вторая, стоящая за буквой переменная y, свободная в исходной формуле, стала обязательной.

Также не приводится подробно описание замены некоторой переменной предиката данной формулой; такая возможность иллюстрируется на таком примере: пусть заменяется переменная предиката F в формуле $(Fxy \wedge Gy) \to Fyy$ (данная формула является двухмерной) формулой $Kxy \wedge Hzx$. Для этого необходимо точно сформулировать, какие переменные заменяющей функции предписываются вместо двух переменных переменной F. Это осуществляется, например, путем подстановки 1 и 2 вместо тех переменных предиката, которые предписываются для первой и второй переменных F.

Некоторые возможные предписания:

$$K21 \wedge Hz2; \qquad Kx2 \wedge H2x; \qquad K1y \wedge H1x; \qquad Kxy \wedge Hzx.$$

Можно видеть, что возможно и такое предписание, которое не содержит обязательства для одной (или обеих) переменных F. Пусть замена проводится, например, по первому предложению. В исходной формуле буква F присутствует дважды. Первое ее наступление имеет вид Fxy; первой переменной здесь является x, а второй — y. Сейчас в заменяющей формуле $K21 \wedge Hz2$, обозначенной цифрами, вместо 1 подставляется x, вместо 2 — y. Полученной таким образом формулой $Kyx \wedge Hzy$ следует заменить формулу Fxy. Второе наступление буквы F имеет вид Fyy. В этом случае в заменяющей формуле с цифрами вместо 1 и 2 подставляется y, следовательно, формула Fyy заменяется формулой $Kyy \wedge Hzy$. Окончательный результат замены имеет вид: $(Kyx \wedge Hzy \wedge Gy) \to (Kyy \wedge Hzy)$.

Если, например, по третьему предложению подставляется выражение $(K1y \wedge H1x)$, то при замене Fxy в заменяющей формуле подставляется x вместо 1 и при замене Fyy — y вместо 2. Таким образом, окончательным результатом замены будет $(Kxy \wedge Hxx \wedge Gy) \to (Kyy \wedge Hyx)$. Но если по четвертому предложению хотелось бы заменить формулу $(Kxy \wedge Hzx)$, то, поскольку в данной заменяющей формуле вообще не фигурируют цифры, каждое наступление буквы F просто заменяется заменяющей формулой; итак, результатом замены будет формула $(Kxy \wedge Hzx \wedge Gy) \to (Kxy \wedge Hzx)$.

Пример запрета при замене переменной предиката. Запрещается, например, в формуле $\forall x\, Fx \to Fy$ замена F формулой $\exists y\, G1y$, так как в полученной таким образом формуле $\forall x\, \exists x\, Gxy \to \exists y\, Gyy$ в последнем члене квантором $\exists y$ заменяющей формулы закрывается наступление такой переменной y, наступление которой было свободным в исходной формуле. (Исходная формула является тавтологией, а полученная путем замены — уже нет!)

АКСИОМАТИЧЕСКОЕ ПОСТРОЕНИЕ КАЛЬКУЛЯЦИИ ПРЕДИКАТОВ

Хотя некоторые формулы представляют собой тавтологии, но на основе заданного определения тавтологии нельзя решать, является ли любая формула тавтологией, так как нет возможности испробовать все оценки (как при калькуляции высказываний). Возникает вопрос: существует ли какой-либо способ распознавания тавтологий?

Аналогично аксиоматическому построению калькуляции высказываний, можно осуществить *аксиоматическое построение калькуляции предикатов*. Одним из таких возможных построений является следующее: пусть тавтологии 1—10, приведенные на стр. 648, дополняются еще и следующими:

11. $\forall x\, Fx \to Fy$ 12. $Fy \to \exists x\, Fx$.

Правилами выведения считаются *замена* (в смысле обобщения для калькуляции предикатов), *отделение* и еще следующие два правила:

а) Если α и β являются формулами, а x — переменной, не наступающей в свободном виде в формуле α, то из формулы $(\alpha) \to (\beta)$ может быть приведена формула $(\alpha) \to \forall x(\beta)$.

б) Если α и β — формулы, а x — переменная, не наступающая в свободном виде в формуле β, то из формулы $(\alpha) \to (\beta)$ может быть приведена формула $\exists x(\alpha) \to (\beta)$.

Можно доказать, что любая тавтология калькуляции предикатов может быть выведена на основе применения тавтологий 1—12 с использованием вышеприведенных четырех правил выведения, учитывая, что по данным правилам тавтологии всегда возникают из тавтологий. (Возможность приведения тавтологий калькуляции предикатов таким путем была доказана Гёделем в так называемой *теореме полноты.)* Это позволяет задавать определение тавтологии независимо от оценки (независимо от понятий теории множеств, присутствующих в определении термина оценки): тавтологиями являются формулы 1—12 («аксиомы»), а, кроме того, все формулы, выводимые из них на основе применения упомянутых выше четырех «правил выведения».

В представленной системе аксиом каждая тавтология может быть выведена. *Зная*, что некоторая формула является тавтологией, ее выведение может быть найдено путем применения ограниченного количества проб (могут быть заданы такие методы, которыми регулируются возможности проб).

Черч доказал, что не всегда возможно нахождение такого общего метода, с помощью которого о любой формуле калькуляции предикатов можно «судить» по результату ограниченного количества шагов (может ли она быть приведена, т. е. является ли она тавтологией или нет). Таким образом, в калькуляции предикатов невозможно существование общего метода решения.

С другой стороны, существует общий метод решения для формул, содержащих переменные предикатов с числом аргументов, не превышающим 1. Из-за ограниченного объема настоящей работы этот метод здесь не представлен.

Применение калькуляции предикатов для исследования правильности выводов осуществляется, по сути дела, таким же образом, как и применение калькуляции высказываний. С другой стороны, более затруднительно распознавание кванторов и предикатов в текстовых высказываниях. Для иллюстрации приводим вывод (рассмотренный во введении настоящей главы):

Предпосылка: Каждая трапеция — четырехугольник или $\forall x$(x трапеция $\rightarrow$ x четырехугольник).

Окончательный вывод: Кто чертит трапецию, тот чертит четырехугольник или

$\forall y[\exists x$(x трапеция $\wedge$ y чертит x) $\rightarrow$ $\exists x$(x четырехугольник $\wedge$ y чертит x)].

Подставляются переменные предикатов вместо предикатов в исходной формуле:

Fx: x трапеция; Gyx: y чертит x; Hx: x четырехугольник.

Таким образом, предпосылке соответствует формула

$\forall x(Fx \rightarrow Hx)$, а окончательному выводу — формула

$$\forall y[\exists x(Fx \wedge Gyx) \rightarrow \exists x(Hx \wedge Gyx)].$$

(Оценка последней формулы была приведена на стр. 655.)

Контролируется правильность схемы вывода

$$\forall x(Fx \rightarrow Hx) \vdash \forall y[\exists x(Fx \wedge Gyx) \rightarrow \exists x(Hx \wedge Gyx)],$$

или (что одно и то же) следует показать, что формула

$$\forall x(Fx \rightarrow Hx) \rightarrow \forall y[\exists x(Fx \wedge Gyx) \rightarrow \exists x(Hx \wedge Gyx)]$$

представляет собой тавтологию. По оригинальному (заданному вместе с термином оценки) определению тавтологии, задача может быть решена на основе следующего рассуждения.

Предположим, что существует такая оценка, для которой данная формула принимает значение *л*. Это означало бы, что действительны следующие соотношения, если F, G, и H интерпретируются предикатами $\mathbf{F}x$, $\mathfrak{G}yx$ и $\mathbf{H}x$ на некотором непустом множестве A:

(1) $$\forall x(\mathbf{F}x \rightarrow \mathbf{H}x) = n \quad \text{и}$$

(2) $$\forall y[\exists x(\mathbf{F}x \wedge \mathfrak{G}yx) \rightarrow \exists x(\mathbf{H}x \wedge \mathfrak{G}yx)] = л.$$

Но тогда из (1) вытекает, что для всех x:

(3) $$\mathbf{F}x \rightarrow \mathbf{H}x = n$$

и на основе соотношения (2) хотя бы для одного элемента y_1 множества A было действительно соотношение

(4) $$\exists x(\mathbf{F}x \wedge \mathbf{G}y_1x) \to \exists x(\mathbf{H}x \wedge \mathbf{G}y_1x) = л.$$

А из-за правильности (4) получилось бы соотношение

(5) $$\exists x(\mathbf{F}x \wedge \mathbf{G}y_1x) = n \quad \text{и}$$

(6) $$\exists x(\mathbf{H}x \wedge \mathbf{G}y_1x) = л.$$

По соотношению (5) хотя бы для одного элемента x_1 множества A действительно соотношение

$$\mathbf{F}x_1 \wedge \mathbf{G}y_1x_1 = n, \quad \text{т. е.}$$

(7) $$\mathbf{F}x_1 = n \quad \text{и}$$

(8) $$\mathbf{G}y_1x_1 = n.$$

Но в таком случае из-за соотношений (7) и (3) было бы $\mathbf{H}x_1 = n$ и, таким образом, учитывая соотношение (8):

$$\mathbf{H}x_1 \wedge \mathbf{G}y_1x_1 = n,$$

но это противоречит условию (6). Следовательно, такая оценка исключается: результат любой оценки вышеприведенной формулы не может быть $л$. Данная формула действительно является тавтологией.

СИЛЛОГИЧЕСКИЕ ВЫВОДЫ

В эпоху средневековья в так называемых *силлогических* выводах, исследованных особенно часто в логике, фигурировали так называемые *категоричные высказывания**. Категоричными высказываниями назывались высказывания, структура которых определялась некоторым видом формул

$$\forall x(Fx \to Gx); \qquad \forall x(Fx \to \neg Gx),$$

$$\exists x(Fx \wedge Gx), \qquad \exists x(Fx \wedge \neg Gx).$$

В некотором силлогическом выводе имеются две предпосылки. Как предпосылки, так и окончательный вывод представляют собой категоричные высказывания; в их формализованном виде присутствуют всего три переменных предиката, каждая из которых присутствует в двух формулах (или в двух предпосылках, или в одной из предпосылок и в окончательном выводе). Нетрудно подсчитать, что число существенно отличающихся

* Точнее говоря, здесь приведены так называемые *категоричные* силлогизмы, в отличие от так называемых *гипотетических* силлогизмов. Последнее название имели некоторые схемы вывода калькуляции высказываний в области традиционной логики.

друг от друга схем, соответствующих таким условиям, составляет 144, хотя, конечно, они не все представляют собой правильные схемы вывода.

Два примера:

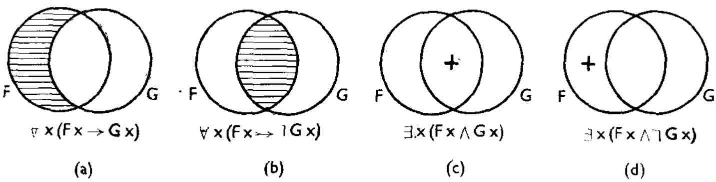
(9) $\forall x(Fx \to Gx)$; $\forall x(Gx \to Hx) \vdash \forall x(Fx \to Hx)$.

(10) $\forall x(Fx \to Gx)$; $\forall x(Fx \to Hx) \vdash \exists x(Gx \wedge Hx)$.

Первая схема является правильной, вторая — неправильной (изложение причин см. ниже).

Для исследования силлогических выводов применяется простой графический метод, так называемый, метод *диаграмм Венна* (см. гл. «Теория множеств»). При использовании данного метода структура некоторого категоричного высказывания изображается двумя кругами, пересекающими друг друга на основании следующих правил.

На рисунках штриховкой обозначается пустота, а крестиком — существование. На рисунке (a) изображена такая ситуация, что F не имеет элемента, не принадлежащего G; т. е. *каждое F заодно — G*: $\forall x(Fx \to Gx)$.

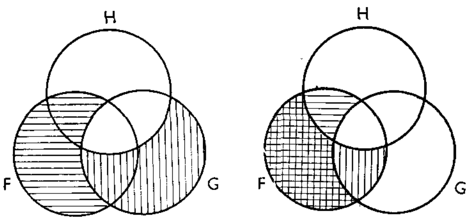


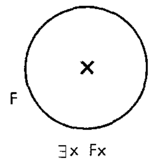
На рисунке (b) F и G не имеют общих элементов: *то, что — F, то — не G*, следовательно: $\forall x(Fx \to \neg Gx)$. По *рисунку* (c) F и G имеют общий элемент, т. е. *существует такой предмет, который — и F, и G*: $\exists x(Fx \wedge Gx)$. Наконец, на графике (d) показана такая ситуация, по которой F имеет элемент, не принадлежащий G, т. е. есть *такой предмет, который — F, но — не G*: $\exists x(Fx \wedge \neg Gx)$. На данных рисунках переменные предикатов F, G представлены кругами; кругами характеризуется *объем* предикатов, соответствующих буквам (класс таких предметов, в отношении которых предикат является правильным). Правильность некоторой схемы вывода контролируется с помощью диаграммы Венна. Начертим три круга, пересекающих друг друга таким образом, чтобы они имели общую часть; переменные предикатов в данной схеме обозначаются буквами F, G, H. В соответствии с указаниями на первом рисунке изображается диаграмма двух предпосылок. Если при этом автоматически создается и диаграмма окончательного вывода, то схема вывода правильна, в противном случае — неправильна.

График схемы (9) готовится следующим образом:

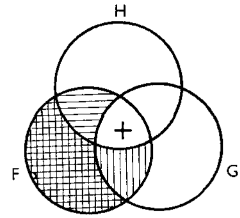
В соответствии с первой предпосылкой зачерчивается часть F, расположенная вне G; затем, в соответствии со второй предпосылкой, зачерчивается часть G, расположенная вне H. Оказывается не заштрихованной часть F, расположенная внутри H. Следовательно, *каждое F — заодно H*:

$\forall x(Fx \to Hx)$. Диаграммой доказывается правильность окончательного вывода. (Данной схемой вывода описывается, например, вывод, представленный на стр. 647, не доказуемый в области калькуляции высказываний.) График схемы (10) представлен ниже. Окончательный вывод схемы (10) имеет следующий вид: $\exists x(Gx \wedge Hx)$. На рисунке на предыдущей странице график (c) соответствует такому положению (но там приведены





(a)



(b)

буквы F и G вместо букв G и H). По графику в данной ситуации в общей части G и H должен был бы находиться крестик. На графике его нет, поэтому данная схема вывода неправильна. Конкретный пример:

Каждый, читающий настоящую книгу, читает и эту страницу.
Каждый, читающий настояшую книгу, читает и следующую страницу.

Имеется такой человек, который читает и эту и следующую страницу.

Две предпосылки наверняка будут правильны. Даже в том случае, если никто не читает настоящей книги. На тогда окончательный вывод — неправильный. Вывод будет правильным только тогда, если прибавить следующую «предпосылку существования»:

Существует такой человек, который читает настоящую книгу.

В общем случае схема (10) может быть исправлена дополнительной предпосыкой $\exists x\, Fx$. Отдельный график для такого случая представлен на рисунке (a). Если и на предыдущем графике изобразить дополнительную предпосылку, то крестик может быть поставлен только в незаштрихованную часть круга F. Модифицированный график показан на рисунке (b): в этом случае диаграммой уже доказывается правильность окончательного вывода.

Среди силлогических выводов имеются 15 правильных и 9 таких, которые могут быть исправлены с помощью одной или двух «дополнительных предпосылок существования». В средние века в логике для распознавания правильности силлогических выводов были разработаны искусственные правила. В результате метода диаграмм Венна, с помощью которого любой силлогический вывод контролируется быстро и просто, такие правила стали совсем безынтересными. Впрочем, метод диаграмм приемлем для исследования правильности не только силлогических схем вывода, но и схем, распространенных в существенно более широком кругу.

Приведенный ниже пример вывода разветвляется на цепь многоступенчатых схем силлогических выводов и контролируется с помощью метода диаграмм.

Предпосылки: Ни одна женщина не получила чека.
Тот, кто заплатил, получил чек.
Каждый гость пил чай.
Кто не член клуба, тот — гость.
Каждый член клуба заплатил.

Окончательный вывод: Каждая женщина пила чай.

Для формализации вывода введутся следующие переменные предикатов вместо приведенных предикатов:

«женщина» — N; «получил чек» — B;

«заплатил» — F; «гость» — V;

«пил чай» — U; «член клуба» — K.

С помощью таких обозначений схема выводов приобретает вид:

$$\begin{array}{lll} \text{предпосылки:} & (1) & \left\{\begin{array}{l} \forall x(Nx \to \neg Bx) \\ \forall x(Fx \to Bx) \\ \forall x(Vx \to Ux) \\ \forall x(\neg Kx \to Vx) \\ \forall x(Kx \to Fx) \end{array}\right. \\ & \begin{array}{r}(2)\\(3)\\(4)\\(5)\end{array} & \\ \hline \text{окончательный вывод:} & (6) & \forall x(Nx \to Ux) \end{array}$$

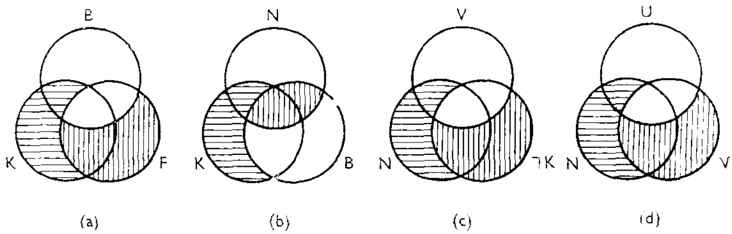
Правильность вывода контролируется на следующих этапах. Вначале приводится силлогический вывод из предпосылок (5) и (2):

$$\text{(a)} \qquad \frac{\forall x(Kx \to Fx);\ \forall x(Fx \to Bx)}{(7)\ \forall x(Kx \to Bx)}$$

Правильность вывода (a) доказывается диаграммой (a) приложенного рисунка. На рисунке приведены графики, доказываемые правильность последующих шагов.

Вывод продолжается на основе полученного таким образом окончательного вывода (7) и предпосылки (1):

(b) $$\frac{\forall x(Kx \to Bx);\ \forall x(Nx \to \neg Bx)}{(8)\ \forall x(Nx \to \neg Kx)}$$



К этому окончательному выводу добавляется предпосылка (4):

(c) $$\frac{\forall x(Nx \to \neg Kx);\ \forall x(\neg Kx \to Vx)}{(9)\ \forall x(Nx \to Vx)}$$

На графике вывода (c) одним кругом представляется символ предиката $\neg K$ (вместо K); таким приемом облегчается применение метода.

Вывод продолжается с помощью вывода (9) и предпосылки (3):

(d) $$\frac{\forall x(Nx \to Vx);\ \forall x(Vx \to Ux)}{(6)\ \forall x(Nx \to Ux)}$$

Результат представляет собой искомый окончательный вывод.

ТОЖДЕСТВО

Среди предикатов *тождество* имеет особое значение. Соотношение, в котором «x тождественно y», обычно обозначается символом «$x=y$». Если предикат тождества используется в некотором выводе, то правильность вывода обычно не удается доказать применением способов калькуляции предикатов. Например, очевидна правильность соотношения $\forall x(x = y \to y = x)$, но соответствующая данному соотношению формула $\forall x(Fxy \to Fyx)$ не является тавтологией. Данный недостаток устраняется, если при формализации предикат тождества не заменяется переменной предиката, а остается в исходном виде. Определение формулы (стр. 654) дополняется следующим условием:

(e) Если x и y — переменные индивидуума, то выражение $x = y$ представляет собой формулу.

Система аксиом дополняется следующими аксиомами:

13. $\forall x(x = x)$ 14. $\forall y\, \forall x\big((x = y \wedge Fx) \to Fy\big)$.

Аксиома 13 выражает, что каждый предмет тождествен самому себе, а аксиома 14 — что если что-либо правильно для одной стороны тождества, то это правильно и для другой стороны (т. е. что тождественные предметы неразличимы). Из этих двух тождеств вытекает, что тождество является симметричным и транзитивным:

$$\boxed{x = y \leftrightarrow y = x \quad \text{и} \quad (x = y \wedge y = z) \rightarrow x = z\,.}$$

Такая модификация калькуляции предикатов называется *калькуляцией предикатов, дополненной тождеством.*

С помощью понятия тождества может быть осуществлена формализация высказывания «по крайней мере существует одно такое x, при котором $\mathbf{F}x$» в отношении некоторого предиката $\mathbf{F}x$:

$$\forall x\, \forall y\big((Fx \wedge Fy) \rightarrow x = y\big),$$

и формализация высказывания «существует именно одно такое x, при котором $\mathbf{F}x$»:

$$\exists x\big(Fx \wedge \forall y(Fy \rightarrow x = y)\big)\,.$$

Последняя формализация иногда сокращается с помощью символа «$\exists!\, x\, Fx$»; символ «$\exists!$» представляет собой так называемый *квантор существования и единственности.*

В области калькуляции предикатов анализируются также так называемые *единственные выводы*, содержащие и такое высказывание в предпосылках или в окончательном выводе, которое возникло в результате замены одной или нескольких переменных одного предиката названием конкретного предмета *(единственного термина).* Например, из предиката «x — неделимое число» возникает высказывание «7 — неделимое число» путем замены буквы x числом 7; или из предиката «$x \geq y$» получится высказывание «$\forall x(x \geq 0)$» путем замены буквы y числом 0 и применения универсальной квантификации x (которое правильно, если, например, x может меняться в множестве натуральных чисел). При формализации единственные термины представляются с помощью переменных индивидуума, а затем вывод контролируется так же, как и в предыдущих случаях. Но в ходе выведения правильности такие переменные не должны присутствовать при использовании правил выведения (а) и (б) в роли x, представленной на стр. 658.

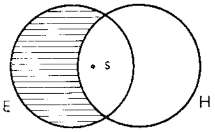
С помощью данного метода контролируется и правильность классического примера традиционной логики:

Предпосылки:	Все люди смертны.
	Сократ человек.
Окончательный вывод:	Сократ смертный.

Обозначения: «человек» — E; «смертный» — H; «Сократ» — s. Соответствующая схема имеет вид:

$$\frac{\forall x(Ex \to Hx)\,; \qquad Es}{Hs}$$

Первая предпосылка изображается обычным путем с помощью диаграммы Венна. Вторая предпосылка может быть изображена таким образом, что внутри круга E записывается точка со знаком s. Диаграмма доказывает правильность окончательного вывода, так как точка находится одновременно и внутри круга H.



В области калькуляции предикатов, дополненной тождеством, единственные выводы могут быть представлены и по-другому. Если a представляет собой единственный термин, то пусть создается предикат $\mathbf{F}x$, являющийся правильным для a и только для a (это — предикат: «быть a»; например, если a = гора Геллерт, то данный предикат: «быть горой Геллерт»). Правильность любого предиката $\mathbf{G}x$ для a может быть представлена в виде:

$$\exists y[\mathbf{G}y \wedge \forall x(\mathbf{F}x \leftrightarrow x = y)]\,.$$

С помощью такого приема высказывания, относящиеся к единственным терминам, могут быть формализованы такими же способами, что и высказывания с кванторами без такого условия, что свободные переменные должны представляться как единственные термины.

В области калькуляции предикатов могут быть формализованы высказывания, относящиеся к предикатам. Множество H, элементы которого заменяются вместо переменных некоторого предиката, может быть множеством или высказываний, или предикатов. Но данная теория оказывается неудовлетворительной, например, для формализации высказываний типа: Для каждого предиката $F^{(1)}$ существует такой предикат $G^{(1)}$, при котором $\forall x(Fx \leftrightarrow Gx)$.

Формализация таких и аналогичных высказываний и анализ выводов, проведенных с их помощью, относится к области *дополненной калькуляции предикатов* (или *калькуляции предикатов более высокого порядка).* (Представленную здесь калькуляцию предикатов для различимости называют *более узкой калькуляцией предикатов или калькуляцией предикатов первого порядка.)* Ограниченный объем настоящей книги не позволяет остановиться на рассмотрении дополненной калькуляции предикатов.

ПРИЛОЖЕНИЯ

МАТЕМАТИЧЕСКИЕ ПРИЛОЖЕНИЯ

Дополненная тождеством (в более узком смысле) калькуляция предикатов преимущественно применяется в анализе структуры математических выведений и для уточнения понятия «аксиоматического построения» (см. гл. «Теория множеств»). Аксиомы некоторой, в достаточной степени точно сформулированной системы аксиом могут быть формализованы в области калькуляции предикатов: основные понятия в аксиомах выражаются переменными предикатов, а логические понятия — логическими знаками ($\neg$, $\wedge$, $\vee$, $\rightarrow$, $\leftrightarrow$, $\forall$, $\exists$, $=$). *Теоремы* системы аксиом выражаются замкнутыми формулами, являющимися *следствиями* формул, выражающих аксиомы.

Система аксиом является непротиворечивой, если не существует формулы α, при которой и α, и $\neg\alpha$ являются следствиями формул, характеризующих аксиомы. Собственными формулами системы аксиом являются такие формулы, в которых нет других переменных предикатов, как в формулах, описывающих аксиомы. Система аксиом называется полной, если для любой ее замкнутой, собственной формулы α формула α или формула $\neg\alpha$ выражает некоторую теорему системы аксиом.

Гёдель доказал, что системы, принадлежащие к обширному классу систем математических аксиом (и именно основные, «стоющие» системы аксиом), не могут быть полными, если они являются непротиворечивыми (и не могут быть дополнены таким образом, чтобы путем дополнения они стали полными). По существу, этим утверждением выражается, что в некоторых областях математики нет универсального метода выведения. (Для сравнения: как указалось выше, в области калькуляции предикатов нет универсального метода решения, но есть универсальный «метод выведения», поскольку каждая тавтология может быть выведена из ограниченного числа тавтологий.)

Если для некоторой совокупности формул калькуляции предикатов в некотором непустом множестве может быть задана интерпретация переменных предикатов, присутствующих в формулах, такая, что при данной интерпретации каждая формула, принадлежащая к данной совокупности, принимает значение *п*, то такую интерпретацию можно осуществить уже в множестве *натуральных чисел*. Это правильно и в том случае, если данной совокупности принадлежит исчислимо бесконечное количество формул (*теорема Лёвенгейма—Сколема*). Например, формулы, характеризующие систему аксиом вещественных чисел или теории множеств, «правдиво интерпретируются» уже и во множестве натуральных чисел. Данная теорема выражает тот факт, что обычно логической структурой аксиом не предписывается, во множестве какой численности аксиомы могут быть правильны.

ТЕХНИЧЕСКИЕ ПРИЛОЖЕНИЯ

Математическая логика может быть применена и в области проектирования электрических цепей и некоторого их упрощения.

Пусть *ключом* называется каждый инструмент, имеющий два альтернативных состояния с точки зрения проводимости электрического тока: токопроводящее и изолирующее. Два ключа называются *синхронными*, если они одновременно проводят или изолируют ток. Два ключа называются асинхронными, если в то время как один из них проводит ток, другой изолирует, и наоборот.

Последовательно-параллельной электрической цепью называется любое из следующих: (а) любой ключ; (б) последовательное включение двух (уже существующих) электрических цепей; (в) параллельное соединение двух цепей. В такой электрической цепи могут быть включены и синхронные, и асинхронные ключи. Сокращенное обозначение таких цепей: *ППЭЦ*.

Электрическая цепь, возникающая в результате последовательного соединения двух электрических цепей, проводит ток тогда и только тогда, если оба ее компонента проводят ток. Электрическая цепь, возникающая в результате параллельного соединения двух электрических цепей, изолирует ток тогда и только тогда, если оба ее компонента изолируют ток.

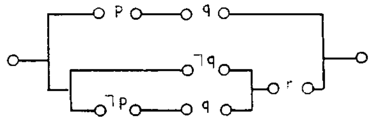
Пусть логические переменные $(p, q, r, \ldots)$ и их отрицания $(\neg p, \neg q, \neg r, \ldots)$ называются *логическими буквами*; p и $\neg p$; q и $\neg q$ и т. д. — противоречием друг другу.

Ключам ППЭЦ логические буквы предписываются по следующему правилу: (а) Синхронным ключам предписываются однинаковые буквы, асинхронным — буквы, противоречивые друг другу. (б) Если два ключа не являются ни синхронными ни асинхронными, то буквы, предписанные им, не должны быть ни одинаковыми, ни буквами, противоречивыми друг другу.

Для каждой ППЭЦ предписывается одна формула калькуляции высказываний по следующим правилам: (а) Ключам предписываются логические буквы в соответствии с правилами, приведенными в предыдущем абзаце. (б) Если к некоторой электрической цепи предписывается формула α, а к другой — формула β, то к электрической цепи, возникающей путем последовательного и параллельного соединения двух данных электрических цепей, предписываются формулы $(\alpha) \wedge (\beta)$ и $(\alpha) \vee (\beta)$ соответственно.

Пусть α представляет собой формулу, предписанную к некоторой ППЭЦ. Некоторая оценка формулы в отношении данного состояния электрической цепи называется *адекватной*, если в электрической цепи проводят ток именно те ключи, предписанные к которым логические буквы имеют значение n в данной оценке. Очевидно, что некоторая ППЭЦ проводит ток тогда и только тогда, если адекватная оценка предписанной к ней формулы в результате дает значение n.

На рисунке представлена схема ППЭЦ. Ключи обозначены прерываниями, кончающимися кругами небольшого радиуса. В ключах записаны

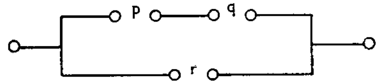


$(p \wedge q) \vee [(\neg q \vee (\neg p \wedge q)) \wedge r]$

предписанные им логические буквы. Под схемой приведена формула, относящаяся к данной электрической цепи.

С помощью способов калькуляции высказываний может быть показана равноценность данной формулы следующей формуле:

$$(p \wedge q) \vee r.$$



$(p \wedge q) \vee r$

Таким образом, схема ППЭЦ, представленная на нижнем рисунке, которой принадлежит настоящая формула, является равноценной предыдущей схеме с точки зрения проводимости тока, но она более простая, экономичная. С помощью соотношений калькуляции высказываний могут быть решены многие задачи упрощения схем электрических цепей. Разработаны специальные методы, служащие целям упрощения формул, описывающих режим электрических цепей, а также методы для проектирования и упрощения не только схем ППЭЦ, но и других электрических цепей (содержащих более, чем два свободных конца, так называемых многополюсных электрических цепей и мостовых электрических схем).

При переключении пар асинхронных ключей оба члена пары изолируют ток в течение некоторого времени. За это короткое время нельзя дать адекватной оценки для формулы, предписанной к данной электрической цепи (так как p и $\neg p$ не могут быть одновременно ложны). Аналогичное явление наступает, если ключи приводятся в движение с помощью релейных схем (электромагнитных реле) и обеспечение током того или другого реле зависит от состояния проводимости тока электрической схемы в данный момент времени. В таких и подобных случаях — и вообще, когда состояние проводимости тока ключей меняется через короткие интервалы времени — для описания состояния проводимости тока электрической цепи в более точной, «адекватной» форме можно прибегать к способам калькуляции предикатов. При проектировании и описании режима функционирования быстродействующих электронных вычислительных автоматов применение калькуляции предикатов является необходимым.

КРАТКИЙ ОБЗОР ИСТОРИИ МАТЕМАТИЧЕСКОЙ ЛОГИКИ

Основы логики и основы математической логики были заложены Аристотелем в IV веке до нашей эры. В течение двух тысячелетий считалось, что его открытие настолько совершенно, что не может иметь дальнейшего развития. Аристотель ввел понятие переменной в логике и применил буквы для обозначения понятий; его последователи в средневековье также внесли большой вклад в дело развития логики, но их результаты не нашли широкого применения из-за слабого развития математики. Г. В. Лейбницем (1646—1716) была выдвинута идея усовершенствования логики. Ему не исполнилось и 20 лет, когда в «Dissertatio de arte combinatoria» он высказался за необходимость применения универсального научного языка. Для этого было необходимым развитие универсального метода мышления, с помощью которого могут быть решены все проблемы, выражаемые с помощью универсального языка. Но своих утопических научных планов он не осуществил.

Следующей важной вехой по пути развития математической логики было выступление Дж. Буля (1815—1864); им обоснована новая алгебраическая система. В его алгебре переменные означают классы, а операции «произведения» и «сложения» создают новые классы путем комбинирования классов. Свои результаты он опубликовал в работах «The mathematical Analysis of Logic» (1847) и затем «The laws of Thought» («Математический анализ логики» и «Правила мышления»). Буль применил свою алгебру в отдельных областях логики, в частности и к силлогизмам классической логики.

А. де Морганом (1806—1871) и У. С. Джевонсом (1835—1882) была развита и упрощена алгебра Буля. Американским ученым Ч. С. Пейрсом (1839—1914) были достигнуты существенные результаты почти во всех областях логики, но действительную цену его результатам узнали только в XX веке, когда интерес целого ряда математиков был привлечен к математической логике. Г. Фреге (1848—1925), Дж. Пеано (1858—1932), Б. Расселль (1872—1970). В работе Б. Расселля в соавторстве с А. Н. Уайтхедом «Principia Mathematica» (1910) была разработана уже полная математическая логическая система, на которой предполагалось построение всей математики.

Деятельность Д. Гильберта (1862—1943) и его учеников сыграла решающую роль в развитии математической логики. Среди его учеников следует упомянуть об известном математике Иоганне фон Неймане.

В настоящее время существенными вопросами математической логики занимаются советская, польская, и американская школы.

Среди известных представителей советской школы по проблемам математической логики много новых результатов было достигнуто П. С. Новиковым, а по прикладным проблемам данной дисциплины — М. А. Гавриловым и В. М. Глушковым.

В Венгрии лучшими представителями математической логики являются Л. Калмар, Р. Петер и Я. Шурани.

ИМЕННОЙ И ПРЕДМЕТНЫЙ УКАЗАТЕЛЬ

В

Д

К

Л

М

П

Р

Т

У

Ф

Ш

Э

Я