НАЧАЛЬНОЕ И СРЕДНЕЕ
ПРОФЕССИОНАЛЬНОЕ
ОБРАЗОВАНИЕ

ФГОС 3+

М.И. Башмаков

МАТЕМАТИКА

Рекомендовано ФГУ «ФИРО»
в качестве **учебника** для использования в учебном процессе образовательных учреждений начального и среднего профессионального образования, реализующих программы среднего (полного) общего образования в пределах основных профессиональных образовательных программ НПО/СПО с учетом профиля получаемого профессионального образования

Министерство образования и науки Российской Федерации
ФГУ «Федеральный институт развития образования»
Регистрационный номер рецензии № 168 от 17.06.2011

BOOK.ru
ЭЛЕКТРОННО-БИБЛИОТЕЧНАЯ СИСТЕМА

КНОРУС • МОСКВА • 2017

УДК 51(075.32)
ББК 22.1я723
Б33

Рецензент
Л.Г. Филимонова, преподаватель математики ФГОУ Санкт-Петербургского техникума железнодорожного транспорта

Башмаков М.И.
Б33 Математика : учебник / М.И. Башмаков. — М. : КНОРУС, 2017. — 394 с. — (Начальное и среднее профессиональное образование).

ISBN 978-5-406-05386-7

Написан в соответствии с программой изучения математики в учреждениях начального и среднего профессионального образования. Охватывает все основные темы программы: теорию чисел, корни, степени, логарифмы, прямые и плоскости, пространственные тела, а также основы тригонометрии, анализа, комбинаторики и теории вероятностей. Алгебра, геометрия и начала анализа излагаются как один учебный предмет. Позволяет успешно подготовиться к итоговой аттестации.

Соответствует ФГОС НПО и СПО 3+.

Для учащихся любых учебных заведений, реализующих программы общего среднего образования.

УДК 51(075.32)
ББК 22.1я723

Башмаков Марк Иванович
МАТЕМАТИКА

Сертификат соответствия № РОСС RU.АГ51.Н03820 от 08.09.2015.

Изд. № 11132. Формат 60×90/16.
Гарнитура «NewtonC». Печать офсетная.
Усл. печ. л. 25,0. Уч.-изд. л. 13,65.

ООО «Издательство «КноРус».
117218, г. Москва, ул. Кедрова, д. 14, корп. 2.
Тел.: 8-495-741-46-28.
E-mail: office@knorus.ru http://www.knorus.ru

Отпечатано в ООО «Контакт».
107150, г. Москва, проезд Подбельского 4-й, д. 3.

ISBN 978-5-406-05386-7

ОГЛАВЛЕНИЕ

ГЛАВА 1

ЧИСЛА, ФУНКЦИИ И ГРАФИКИ

§ 1. Числовая прямая

Целые и рациональные числа

Число — основное понятие математики. Счет предметов осуществляется с помощью *натуральных чисел*. Русское слово «пять» обозначает количество предметов в любом множестве, в котором этих предметов столько же, сколько, например, пальцев на одной руке человека. Арабская цифра 5 служит принятым обозначением этого числа. Натуральные числа выстроены в последовательность, которая начинается с единицы и где каждое следующее число на единицу больше предыдущего.

Арабские цифры 1, 2, 3, 4, 5, 6, 7, 8, 9 служат для записи первых натуральных чисел. Любое натуральное число можно записать в *десятичной системе счисления*, используя эти цифры и еще цифру 0.

Множество всех натуральных чисел обычно обозначается через $\mathbb{N}$.

Целые числа получаются из натуральных добавлением нуля и чисел, обозначаемых: -1; -2; -3; ... Целые числа можно расположить в виде последовательности, бесконечной в обе стороны: ...; -5; -4; -3; -2; -1; 0; 1; 2; 3; 4; 5; ...

Слева от нуля располагаются *отрицательные целые числа*, справа — *положительные*. Положительные целые числа — это те же натуральные числа.

Множество всех целых чисел обычно обозначается через $\mathbb{Z}$.

Отношения натуральных чисел называются положительными *рациональными* числами. Их можно записывать дробями вида $\dfrac{m}{n}$, где m и n — натуральные числа. При этом надо иметь в виду, что две дроби $\dfrac{m}{n}$ и $\dfrac{p}{q}$ являются записями одного и того же рационального числа тогда и только тогда, когда равны (совпадают) натуральные числа mq и pn, т.е.

$$\frac{m}{n} = \frac{p}{q} \Leftrightarrow mq = pn.$$

Логический значок $\Leftrightarrow$ мы будем использовать для записи эквивалентных, равносильных высказываний, т.е. высказываний, которые одновременно верны или нет (истинны или ложны).

Заметим, что натуральное число n можно записать в виде дроби $\frac{n}{1}$. Снабжая дроби знаком минус, мы получим отрицательные рациональные числа. Положительные и отрицательные рациональные числа вместе с нулем образуют множество всех *рациональных чисел*, которое обычно обозначается через $\mathbb{Q}$.

Дальнейшее развитие этих понятий было связано с потребностями измерений.

Что означает измерить длину некоторого отрезка?

Прежде всего, это означает, что мы хотим его сравнить с другим отрезком, выбранным в качестве единицы измерения. Единица измерения может иметь собственное имя, название, скажем, метр, дюйм, аршин, парсек и др. Тогда при записи результата измерения обычно указывают, в каких единицах оно произведено. Мы этого делать не будем и припишем единичному отрезку длину 1 без указаний размерности.

Если единичный отрезок E укладывается в отрезке AB целое число раз, скажем, 3 раза, то натуральное число 3 принимается в качестве длины измеряемого отрезка: $|AB| = 3$.

Пусть отрезок AB соизмерим с единичным, т.е. найдется такой третий отрезок D, который укладывается целое число раз как в единичном отрезке E, так и в отрезке AB. Тогда в качестве длины отрезка AB принимается рациональное число $\frac{m}{n}$, где натуральные числа m и n указывают, сколько раз отрезок D (общая мера) уложится в отрезках AB и E.

Действительные числа

Важнейшим открытием античной математики было обнаружение несоизмеримых отрезков. Так, диагональ квадрата оказалась несоизмерима с его стороной. Геометрическое доказательство этого факта принадлежит Евклиду. Пифагору приписывается другое его доказательство. Если воспользоваться теоремой Пифагора, то квадрат длины диагонали единичного квадрата должен быть равен двум. Иррациональность числа $\sqrt{2}$ и означает несоизмеримость диагонали квадрата с его стороной.

Доказательство иррациональности $\sqrt{2}$. Пусть найдется несократимая дробь $\frac{m}{n}$ такая, что $\left(\frac{m}{n}\right)^2 = 2$. Тогда $m^2 = 2n^2$. Отсюда следует, что число m четно, т.е. $m = 2k$. Получаем $4k^2 = 2n^2$, $2k^2 = n^2$. Отсюда следует, что число n должно быть четным, но это противоречит несократимости дроби.

Длину отрезка, соизмеримого с единичным, можно найти за конечное число шагов с помощью некоторого алгоритма, известного еще Евклиду.

Длину отрезка, несоизмеримого с единичным, можно определить, используя бесконечные алгоритмы. На каждом шагу такого алгоритма будет получаться *рациональное число*, считающееся приближенным значением искомой длины. Последовательность таких рациональных приближений и можно принять за новое число.

Число $\sqrt{2}$, т.е. длину диагонали единичного квадрата, можно записать в виде *бесконечной десятичной дроби* 1,4142135... Эта дробь представляет собой краткую запись последовательности рациональных чисел 1; 1,4; 1,41; 1,414; 1,4142; ... Каждое такое число является некоторым рациональным приближением к числу $\sqrt{2}$. Такая запись удобна, и ее обычно принимают в качестве определения *действительного* (вещественного) *числа*.

Действительное число — это число, записываемое бесконечной десятичной дробью.

Конечные десятичные дроби можно отождествить с бесконечными, добавляя в конце нули: 1,4 = 1,4000... Любое положительное рациональное число можно записать в виде бесконечной десятичной дроби. Например, $\frac{4}{3}$ = 1,333... Среди всех бесконечных десятичных дробей те из них, которые являются записями рациональных чисел, выделяются тем, что они *периодичны*. Таким образом, иррациональные числа — это непериодические десятичные дроби. Более подробно вопрос о связи между рациональными числами и бесконечными десятичными дробями будет рассматриваться далее.

Множество *всех действительных чисел* обычно обозначается через $\mathbb{R}$.

Используя знак $\subset$ для обозначения того, что одно множество содержится в другом, можно записать цепочку включений.

$$\mathbb{N} \subset \mathbb{Z} \subset \mathbb{Q} \subset \mathbb{R}.$$

Числовая ось

Прямая с указанной начальной точкой, выбранным положительным направлением отсчета и фиксированным отрезком (единицей масштаба) называется *числовой осью* (рис. 1).

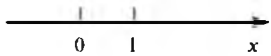


Рис. 1

Каждое число можно изобразить точкой числовой оси.

Наоборот, каждая точка оси изображает какое-нибудь число.

Если число x изображается точкой M, то это число называется *координатой точки M*. Запись $M(x)$ означает: точка M имеет координату x (рис. 2).

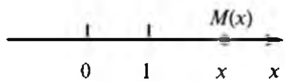


Рис. 2

Соответствие между числами и точками настолько естественно, что часто не различают число и изображающую его точку. Говорят, например, «точка 2», «точка — 0,5», «нулевая точка».

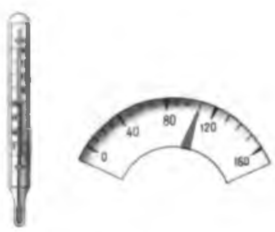


Рис. 3

На шкалы различных измерительных приборов нанесены числа. Этим они похожи на отрезки числовой оси. При этом используют не только прямолинейные шкалы, но и круговые (рис. 3).

Среди числовых множеств особенно часто встречаются множества чисел, лежащих между двумя точками числовой оси. Эти множества называют *числовыми промежутками*. В книгах используют различные термины для их обозначения — отрезок, сегмент, интервал, промежуток. Мы не будем делать различия между этими словами, считая их синонимами. Чаще всего мы будем применять слово «промежуток».

Примечания: 1. Обозначаются промежутки с помощью скобок. При этом квадратная скобка указывает на то, что соответствующий конец промежутка включается в рассматриваемое множество, а круглая — на то, что не включается.

2. Координата начала O равна нулю.

Модуль числа

Модулем числа x называется расстояние от начала отсчета до точки, изображающей число x. Модуль числа x обозначается так: $|x|$.

Модуль числа x можно определить следующим образом:

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

Действительно, возьмем число x и изобразим его точкой M на числовой оси (рис. 4). Если $x \geq 0$, то расстояние $|OM|$ равно x. Если $x < 0$, то это расстояние равно $(-x)$:

$$|OM| = \begin{cases} x, \text{ если } x > 0; \\ -x, \text{ если } x < 0. \end{cases}$$

Так как $|x|$ есть, по определению, расстояние $|OM|$, то мы и получаем требуемую формулу. Она позволяет «раскрывать» модуль, т.е. записывать выражения без знака модуля.

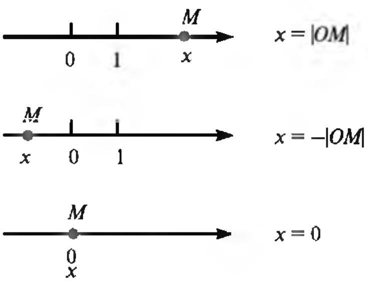


Рис. 4

Теорема. Модуль разности двух чисел равен расстоянию между точками, изображающими эти числа.

Доказательство. Возьмем числа a и b. Обозначим на числовой оси точки, изображающие числа a, b и $a - b$, через A, B и C (рис. 5). При сдвиге вдоль оси x на b точка O перейдет в точку B, а точка C — в A, т.е. $|OC| = |BA|$. Так как, по определению модуля, $|OC| = |a - b|$, то $|a - b| = |BA| = |AB|$, что и требовалось доказать.

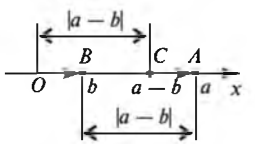


Рис. 5

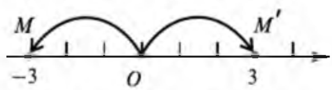
Модуль разности можно раскрыть аналитически:

$$|a - b| = \begin{cases} a - b, \text{ если } a \geq b; \\ b - a, \text{ если } a < b. \end{cases}$$

Простые уравнения и неравенства с модулем удобно решать, используя геометрический его смысл, не раскрывая запись модуля в виде формулы.

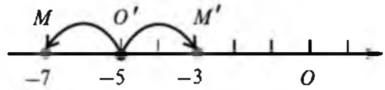
Примеры

1. $|x| = 3$. Это соотношение геометрически означает, что расстояние от точки до начала координат равно 3, т.е. $x = 3$ или $x = -3$ (рис. 6). *Ответ:* $x = \pm 3$ (возможна другая форма записи ответа: $x_1 = -3, x_2 = 3$).
2. $|x + 5| = 2$. Рассматривая $|x + 5|$ как $|x - (-5)|$, прочтем исходное соотношение так: расстояние от точки x до точки -5 равно 2 (рис. 7). Откладывая на числовой оси от точки -5 отрезок длиной 2 (в обе стороны), получаем ответ: $x_1 = -7, x_2 = -3$.



$|OM| = |OM'| = 3$

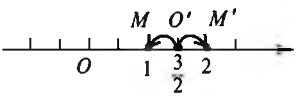
Рис. 6



$|O'M| = |O'M'| = 2$

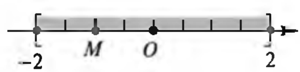
Рис. 7

3. $|3 - 2x| = 1$. Сначала делаем преобразования: $|3 - 2x| = |2x - 3| = \left|2\left(x - \frac{3}{2}\right)\right| = 2\left|x - \frac{3}{2}\right|$. Разделив обе части уравнения $|3 - 2x| = 1$ на 2, находим $\left|x - \frac{3}{2}\right| = \frac{1}{2}$. Используя числовую ось, получаем ответ: $x_1 = 1, x_2 = 2$ (рис. 8).
4. $|x| \le 2$. Задачу решения этого неравенства геометрически можно сформулировать так: найти точки x, расстояние которых до начала координат ≤ 2. *Ответ:* $-2 \le x \le 2$, или в форме промежутка: $[-2, 2]$ (рис. 9).



$|O'M| = |O'M'| = \frac{1}{2}$

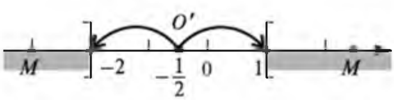
Рис. 8



$|OM| \le 2$

Рис. 9

5. $|-1 - 2x| \ge 3$. Делаем преобразования: $|-1 - 2x| \ge 3 \Leftrightarrow |2x + 1| \ge 3 \Leftrightarrow 2\left|x + \frac{1}{2}\right| \ge 3 \Leftrightarrow \left|x + \frac{1}{2}\right| \ge \frac{3}{2}$ (рис. 10). Ответом является объединение двух бесконечных промежутков $x \le -2$, $x \ge 1$, или другая форма записи: $(-\infty; -2] \cup [1; +\infty)$.



$|O'M| \ge \frac{3}{2}$

Рис. 10

Свойства модуля

1. $|x| = |-x|$.
2. $|x - a| = |a - x|$.
3. $|kx| = |k| \cdot |x|$.
4. $\sqrt{x^2} = |x|$.

Докажем первое свойство. Построим точки M и M' с координатами x и $-x$ соответственно. Так как эти точки симметричны друг другу относительно начала отсчета O, то расстояния $|OM| = |x|$ и $|OM'| = |-x|$ равны, что и требовалось доказать.

Второе свойство является следствием первого, так как $a - x = -(x - a)$. Его можно доказать независимо, используя геометрический смысл модуля разности.

Для доказательства третьего свойства проще всего перебрать все четыре комбинации знаков k и x. Скажем, если $k < 0$, $x > 0$, то $kx < 0$ и $|k| = -k$, $|x| = x$, $|kx| = -kx$, т.е. $|kx| = |k| \cdot |x|$. Остальные случаи разбираются аналогично.

Перед доказательством четвертого свойства необходимо вспомнить определение арифметического корня: $\sqrt{x^2}$ — это такое число z, которое ≥ 0 и квадрат которого равен x^2. Так как $|x| > 0$ и $|x|^2 = x^2$, то $z = |x|$, что и требовалось доказать.

Приближенные вычисления

Мы часто говорим: «От Москвы до Петербурга около 700 километров», «На улице примерно 16 градусов тепла», «Площадь комнаты приблизительно равна 22,6 квадратных метра» и др. Во всех этих случаях идет речь о приближенных значениях некоторых величин — расстояния, температуры, площади.

В приведенных примерах содержится определенная информация, однако часто информации такого сорта недостаточно — приближенному значению можно придать более точный смысл, если привести не только одно число, но и указать точность, с какой оно приближает истинное значение.

Прежде всего для описания точности вычислений применяется термин *погрешность*, который является синонимом слова *ошибка*.

Если точное значение величины равно x, а вычисленное приближенное значение равно a, то *погрешностью вычисления* называется модуль разности точного и приближенного значений, т.е. число $|x - a|$.

Разумеется, если бы мы знали приближенное значение величины и значение погрешности, то мы точно знали бы саму величину, добавив информацию о том, с недостатком или с избытком произведено вычисление. Действительно, если мы знаем величины a и $h = |x - a|$, а также то, что, например, $a > x$, то немедленно нашли бы, что $x = a - h$. То есть геометрически это означает, что искомая величина x отстоит от известного значения a на расстояние h и находится слева от него.

Обычно удается найти не саму погрешность, а *оценку* погрешности, т.е. величину h, для которой верно неравенство $|x - a| \leq h$. Задачу приближенного вычисления некоторой величины обычно так и ставят: найти приближенное значение величины и ***оценить допущенную погрешность.***

Чаще всего в приближенных вычислениях используют *округленные значения* величин в десятичной записи. Так, округленными значениями числа $\pi = 3,1415926536...$ будут

3 — с точностью до 1;
3,1 — с точностью до 0,1;
3,14 — с точностью до 0,01;
3,142 — с точностью до 0,001;
3,1416 — с точностью до 0,0001 и т.д.

Числа 1; 0,1; 0,01; ...дают оценки погрешности для десятичных приближений к числу π. Например, неравенство $|\pi - 3,142| \leq 0,001$ означает, что число 3,142 является приближенным значением числа π, причем погрешность вычисления не превышает 0,001.

Предположим, что, вычисляя приближенные значения двух различных величин, мы в одном случае получили бы в качестве оценки погрешности число $h_1 = 100$, а во втором — $h_2 = 0,01$. Казалось бы, второе вычисление сделано намного точнее, чем первое. Однако это не совсем так. Если первая величина очень велика (например, расстояние от Земли до Солнца в километрах), а вторая очень мала (например, диаметр молекулы в см), то первая оценка окажется очень точной, а вторая — очень грубой. Чтобы различить такие случаи, вводят понятие ***относительной погрешности.***

Пусть a является приближенным значением величины x, вычисленным с погрешностью h, т.е. пусть $|x - a| = h$. Отношение погрешности к приближенному значению, т.е. величину $r = \frac{h}{a} = \frac{|x-a|}{a}$ называют ***относительной погрешностью*** вычисления.

Так, если среднее расстояние от Земли до Солнца вычислено приближенно как $1{,}496 \cdot 10^8$ км с погрешностью $< 10^5$ км, то относительная погрешность такого вычисления будет меньше числа 0,0007, потому что $\frac{10^5}{1{,}496 \cdot 10^8} < 0{,}0007$.

Часто относительную погрешность (а точнее, ее оценку) указывают в процентах, умножая ее значение на 100. Так, можно сказать, что приведенное вычисление сделано с относительной погрешностью в 0,07%. Мы видим в этом примере, что несмотря на то, что оценка погрешности вычисления расстояния от Земли до Солнца выглядит очень большой (100 000 км), относительная погрешность является достаточно малой (меньше 0,0007).

Чтобы отличить погрешность от относительной погрешности, первую часто называют *абсолютной* погрешностью. Если ясно, о какой погрешности идет речь, то слово *абсолютная* опускают.

Приближенные значения величины часто указывают в так называемой *стандартной записи*. Положительные числа в стандартной записи выглядят так: $a \cdot 10^k$, где величину a выбирают так, чтобы оно лежало в промежутке $[1; 10)$, т.е. удовлетворяло неравенствам $1 \leq a < 10$ и записывалось десятичной дробью с несколькими знаками после запятой. Число a в стандартной записи x называют *мантиссой* числа x, а показатель k — его *порядком*.

Так, радиус Земли (если считать Землю шаром) равен числу $R = 6380$ км, что в стандартной записи выглядит так: $R = 6{,}38 \cdot 10^3$ км. Аналогично, масса электрона запишется так: $m = 9{,}11 \cdot 10^{-28}$ г.

Примеры

Указание точности вычислений может быть сделано по-разному. Опишем несколько самых распространенных способов для оценки погрешности.

1. «*Плюс-минус*». Часто говорят так: «температура равна 16 плюс-минус один градус» и записывают: $t = (16 \pm 1)$ °С. Это означает, что истинное значение температуры (в градусах) отличается от 16 не более, чем на единицу. Эту же информацию можно записать в виде неравенства $16 - 1 < t < 16 + 1$ (градус), или с помощью расстояния: $|t - 16| < 1$. Здесь 16 — приближенное значение температуры, 1 — оценка погрешности. Относительная погрешность равна $\frac{1}{16} = 0{,}0625$, т.е. 6,25%.

2. «С точностью до...». Если вы скажете, что площадь комнаты равна 22,6 с точностью до двух десятых квадратного метра, то всем будет ясно, что площадь S лежит в промежутке $22{,}4 < S < 22{,}8$ м^2 или иначе, что расстояние истинного значения площади до числа 22,6 меньше 0,2, т.е. что $|S - 22{,}6| < 0{,}2$ м^2.
   Мы видим, что этот способ фактически совпадает с первым, и мы можем с таким же успехом записать: $S = 22{,}6 \pm 0{,}2$ м^2 и сказать, что площадь вычислена с оценкой погрешности в 0,2 м^2, что дает относительную погрешность, равную $\dfrac{0{,}2}{22{,}6} \approx 0{,}088$, т.е. 9%.
3. «*Лежит между*». Фраза «скорость автомобиля лежит между 50 и 60 километрами в час» сразу определяет промежуток, где лежит значение скорости v: $50 < v < 60$. Можно, конечно, взять середину этого промежутка и перейти к обсуждавшимся ранее способам записи:

$$v = (55 \pm 5) \text{ км/ч};$$
$$|v - 55| < 5 \text{ км/ч}.$$

   Величина 5 км/ч дает оценку погрешности вычисления скорости, а число $\dfrac{5}{55} \approx 0{,}09$ — оценку для относительной погрешности.

Контрольные вопросы и задания

1. Обратите внимание на встретившиеся в тексте ключевые слова и обозначения: числовая ось, координата точки, числовой промежуток, модуль, $M(x)$, $|x|$. Приведите примеры их использования.
2. Дайте определение числовой оси.
3. Что такое координата точки?
4. Как построить на числовой оси точку, если задана ее координата?
5. Как вычислить координату точки числовой оси?
6. Приведите примеры натуральных, целых, рациональных и иррациональных чисел.
7. Приведите примеры различных числовых промежутков.
8. Что такое модуль числа?
9. Чему равен модуль разности двух чисел?
10. Какие вы знаете свойства модуля?

§ 2. Комплексная плоскость

Определение комплексных чисел

Комплексным числом называется выражение вида $a + bi$, где a и b — произвольные вещественные числа; i — специальный символ.

Для комплексных чисел определены действия сложения и умножения. *Суммой комплексных чисел* $z_1 = a_1 + b_1 i$ и $z_2 = a_2 + b_2 i$ называется комплексное число $z = (a_1 + a_2) + (b_1 + b_2)i$, т.е. при сложении комплексных чисел складываются их вещественные и мнимые части.

Умножение комплексных чисел производится по следующему правилу:

$$(a_1 + b_1 i)(a_2 + b_2 i) = (a_1 a_2 - b_1 b_2) + (a_1 b_2 + a_2 b_1)i.$$

Запишем число $0 + 1i$ как i. По определению умножения $i^2 = (0 + 1i) \times \times (0 + 1i) = (0 - 1) + (0 - 0)i = -1$, поэтому число i называют *мнимой единицей*. Заметим, что квадраты вещественных чисел (именно с ними мы имели дело до сих пор) всегда неотрицательны.

Вещественные числа a и b называют соответственно *вещественной* и *мнимой* частями комплексного числа $z = a + bi$. Комплексное число, у которого мнимая часть равна нулю, т.е. комплексное число вида $a + 0i$ записывают в виде a и отождествляют с вещественным числом a.

Два комплексных числа $z_1 = a_1 + b_1 i$ и $z_2 = a_2 + b_2 i$ считают равными тогда и только тогда, когда равны их вещественные и мнимые части:

$$a_1 + b_1 i = a_2 + b_2 i \Leftrightarrow \begin{cases} a_1 = a_2, \\ b_1 = b_2. \end{cases}$$

Комплексные числа можно поставить во взаимно однозначное соответствие с парами вещественных чисел: $a + bi \leftrightarrow (a; b)$.

Это подсказывает геометрическую интерпретацию комплексных чисел — если мы введем на плоскости систему координат, то можно сопоставить каждому комплексному числу $z = a + bi$ точку плоскости $M(a; b)$ с координатами a и b. Это сопоставление будет взаимно однозначным, т.е. каждой точке плоскости M с координатами (a, b) мы можем сопоставить комплексное число $a + bi$. Так же как вещественное число мы отождествляем с соответствующей точкой числовой оси, мы можем отождествить комплексное число с точкой плоскости. Поэтому можно сказать: рассматривая точку $z = 1 + i$, понимаем под этим точку плоскости с координатами (1; 1). Вся плоскость называется *комплексной плоскостью.*

Вещественные числа, рассматриваемые как часть комплексных чисел, занимают ось абсцисс координатной плоскости (вещественная ось). Числа вида $0 + bi$, которые записывают в виде bi, занимают ось ординат (мнимая ось).

Множество всех комплексных чисел обозначают буквой $\mathbb{C}$.

Арифметические действия над комплексными числами

При *сложении комплексных чисел* выполняются обычные арифметические законы.

1. $z_1 + z_2 = z_2 + z_1$ — коммутативный, или переместительный, закон сложения.
2. $(z_1 + z_2) + z_3 = z_1 + (z_2 + z_3)$ — ассоциативный, или сочетательный, закон сложения.
3. Число $0 = 0 + 0i$ играет роль нуля: $z + 0 = 0 + z = z$.
4. Для числа $z = a + bi$ число $(-z) = -a - bi$ играет роль противоположного числа: $z + (-z) = (-z) + z = 0$.

Эти законы выполняются для вещественных чисел. Их проверка для комплексных чисел выполняется без затруднений, так как при сложении комплексных чисел их вещественные и мнимые части «не перемешиваются».

Заметим, что определенное нами сложение комплексных чисел, примененное к вещественным числам (как части комплексных чисел), совпадает с известным ранее сложением вещественных чисел.

Определение *умножения комплексных чисел* можно дать также следующим образом. При умножении комплексных чисел $z_1 = a_1 + b_1 i$ и $z_2 = a_2 + b_2 i$ мы умножаем каждый член на каждый, разрешаем переставлять сомножители и **заменяем произведение $i \cdot i = i^2$ на вещественное число -1**:

$$(a_1 + b_1 i)(a_2 + b_2 i) = a_1 a_2 + a_1 b_2 i + b_1 i a_2 + b_1 i b_2 i =$$
$$= a_1 a_2 + a_1 b_2 i + a_2 b_1 i + b_1 b_2 i^2 = (a_1 a_2 - b_1 b_2) + (a_1 b_2 + a_2 b_1) i.$$

Как мы видим, произведение комплексных чисел определяется более сложно, чем сложение — вещественные и мнимые части не выступают изолированно друг от друга, а перемешиваются. Поэтому арифметические законы умножения требуют проверки, которая не сложна, но занимает довольно много места, и мы ее пропускаем.

5. $z_1 \cdot z_2 = z_2 \cdot z_1$ — коммутативный закон умножения.

6. $(z_1 \cdot z_2) \cdot z_3 = z_1 \cdot (z_2 \cdot z_3)$ — ассоциативный закон умножения.
7. Число $1 = 1 + 0i$ при умножении на любое комплексное число оставляет это число без изменений:
   $1 \cdot z = (1 + 0i) \cdot (a + bi) = 1 \cdot a - 0 \cdot b + (1 \cdot b + 0 \cdot a)i = a + bi = z$.
8. Для каждого числа $z \neq 0$ существует обратное число.

Укажем формулу для обратного числа, вернувшись к ее происхождению далее, если $z = a + bi \neq 0$, то вещественное число $a^2 + b^2$ отлично от нуля, и можно рассмотреть число $z^{-1} = \dfrac{a}{a^2+b^2} - \dfrac{b}{a^2+b^2}i$. Проверкой убеждаемся, что $z \cdot z^{-1} = 1$:

$$(a+bi)\left(\frac{a}{a^2+b^2} - \frac{b}{a^2+b^2}i\right) = \left(a \cdot \frac{a}{a^2+b^2} + b \cdot \frac{b}{a^2+b^2}\right) +$$

$$+\left(a \cdot \frac{-b}{a^2+b^2} + b \cdot \frac{a}{a^2+b^2}\right)i = \frac{a^2+b^2}{a^2+b^2} + 0 \cdot i = 1.$$

Заметим, что если число z — вещественное, т.е. если $z = a$ и $a \neq 0$, то обратное к z, вычисленное по данному определению, равно $z^{-1} = \dfrac{a}{a^2} + 0 = \dfrac{1}{a} = a^{-1}$, т.е. совпадает с обычным обратным, известным для вещественных чисел.

9. К перечисленным основным законам арифметических действий добавим дистрибутивный, или распределительный закон:

$$z_1(z_2 + z_3) = z_1z_2 + z_1z_3.$$

Число 0 при действии сложения и число 1 при действии умножения называются ***нейтральными*** элементами, так как они не изменяют числа, с которыми взаимодействуют: $0 + z = z + 0 = z$, $1 \cdot z = z \cdot 1 = z$.

Таким образом, множество $\mathbb{C}$ обладает девятью основными законами арифметических действий. Такие законы выполняются для чисел множества $\mathbb{R}$ всех вещественных чисел и множества $\mathbb{Q}$ всех рациональных чисел. В алгебре множество, в котором есть две операции (сложение и умножение), подчиняющиеся указанным девяти законам, называют ***полем***. Удобно говорить ***поле*** комплексных чисел, ***поле*** рациональных чисел вместо бесформенного термина ***множество***.

Заметим, что теория сравнений позволяет построить новые поля: множество классов вычетов (сравнений) по модулю p, где p — простое число, тоже образует поле. Таким образом, для каждого простого числа p можно построить ***поле*** из p элементов. Такие конечные поля носят имя замечательного французского математика Э. Галуа и обозначаются F_p.

Арифметические действия над комплексными числами допускают наглядную геометрическую интерпретацию. Начнем со сложения. Пусть даны два положительных числа $z_1 = a_1 + b_1 i$ и $z_2 = a_2 + b_2 i$. Им соответствуют точки $M_1(a_1;\ b_1)$ и $M_2(a_2;\ b_2)$. Сумма чисел $z = z_1 + z_2 = (a_1 + a_2) + (b_1 + b_2)i$ соответствует точке M, координаты которой равны сумме координат точек M_1 и M_2. Это означает, что сложение комплексных чисел соответствует правилу параллелограмма для сложения векторов: $\overline{OM} = \overline{OM_1} + \overline{OM_2}$.

Геометрическая интерпретация умножения будет рассмотрена далее.

Комплексное сопряжение

Мнимая единица i по определению умножения обладает тем свойством, что ее квадрат равен -1, т.е. она является квадратным корнем из -1. Комплексное число $-i$ обладает тем же свойством: $(-i)^2 = (-1 \times \times i)^2 = (-1)^2 \cdot i^2 = -1$, что, разумеется, не удивительно. Можно сказать, что один квадратный корень из -1 мы обозначим через i, а тогда второй корень запишется как $(-i)$. Замена i на $(-i)$ приводит к понятию комплексного сопряжения.

Комплексно сопряженным с числом $z = a + bi$ называется число $a - bi$, которое обозначается $\bar{z}$.

Таким образом, мы определили отображение, по которому каждому комплексному числу z ставится в соответствие число $\bar{z}$, комплексно сопряженное с ним.

Комплексное сопряжение имеет наглядную геометрическую интерпретацию — точки $M(a;\ b)$ и $\overline{M}(a; -b)$, соответствующие комплексно сопряженным числам

$$z = a + bi \text{ и } \bar{z} = a - bi,$$

симметричны друг другу относительно вещественной оси.

Перечислим свойства этого отображения.

1. $\bar{\bar{z}} = z$, т.е. если мы два раза выполним комплексное сопряжение, то вернемся к исходному числу.
2. $\bar{z} = z \Leftrightarrow z \in \mathbb{R}$, т.е. комплексное число равно своему сопряженному в том и только в том случае, когда оно вещественное (его мнимая часть равна нулю).
3. $\overline{z_1 + z_2} = \bar{z}_1 + \bar{z}_2$, т.е. сопряженное к сумме комплексных чисел равно сумме сопряженных.

4. $\overline{z_1 + z_2} = \bar{z}_1 \cdot \bar{z}_2$, т.е. сопряженное к произведению комплексных чисел равно произведению сопряженных.

Первые три свойства очевидны. Полезно заметить, что они отражают простые свойства осевой симметрии. Из этих свойств лишь последнее не очевидно и нуждается в проверке:

$$\overline{(a_1 + b_1 i)(a_2 + b_2 i)} = \overline{(a_1 a_2 - b_1 b_2 + (a_1 b_2 + a_2 b_1) i)} =$$
$$= a_1 a_2 - b_1 b_2 - (a_1 b_2 + a_2 b_1) i.$$

$\overline{(a_1 + b_1 i)} \cdot \overline{(a_2 + b_2 i)} = (a_1 - b_1 i)(a_2 - b_2 i) = a_1 a_2 - (-b_1) \cdot (-b_2) + (a_1 \cdot (-b_2) + a_2 \cdot (-b_1)) i = a_1 a_2 - b_1 b_2 - (a_1 b_2 + a_2 b_1) i$, что и требовалось доказать.

Вычислим произведение числа z на сопряженное:

$$z \cdot \bar{z} = (a + bi)(a - bi) = a^2 + b^2,$$

т.е. мы видим, что это произведение является вещественным числом. Оно имеет простой геометрический смысл — оно равно квадрату расстояния от точки $M(a; b)$ до начала координат или, как мы условились говорить, от точки z на комплексной плоскости до нуля: $z \cdot \bar{z} = |OM|^2$. Число $\sqrt{z \cdot \bar{z}}$, равное длине отрезка OM, называют *модулем* комплексного числа z, т.е. $|z| = \sqrt{a^2 + b^2}$. Заметим, что если $z = a$, то $|z| = |a|$.

Отметим *свойства модуля*.

1. $|z| \geq 0$, причем $|z| = 0 \Leftrightarrow z = 0$.
2. $|z| = |\bar{z}|$.
3. $|z_1 z_2| = |z_1| \cdot |z_2|$, т.е. модуль произведения комплексных чисел равен произведению модулей.
4. $|z_1 + z_2| \leq |z_1| + |z_2|$.

Первые два из этих свойств очевидны. Доказательство третьего свойства сводится к проверке алгебраического тождества:

$$(a_1 a_2 - b_1 b_2)^2 + (a_1 b_2 + a_2 b_1)^2 = (a_1^2 + b_1^2)(a_2^2 + b_2^2).$$

Свойство 4 — это запись геометрического неравенства треугольника (рис. 11).

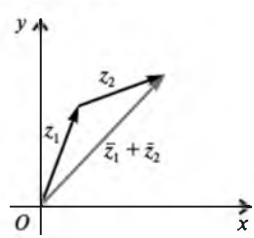


Рис. 11

Формула Кардано

Приведем классический вывод формулы для корней кубического уравнения, в результате которого и появились комплексные числа.

Пусть мы хотим решить уравнение $x^3 + ax^2 + bx + c = 0$. Сначала заметим, что после простой подстановки $x = y - \frac{a}{3}$ оно приводится к уравнению вида $y^3 + py + q = 0$, в котором отсутствует член с квадратом неизвестной. Геометрически это соответствует тому, что мы хотим сделать сумму корней (которая равна коэффициенту при квадрате с противоположным знаком) равной нулю за счет изменения начала отсчета. Алгебраически это сводится к проверке того, что члены с величиной y^2 взаимно уничтожаются: $x^3 + ax^2 = \left(y - \frac{a}{3}\right)^3 + a\left(y - \frac{a}{3}\right)^2 =$ $= y^3 - \underline{3y^2 \cdot \frac{a}{3}} + \ldots + \underline{ay^2} + \ldots$ Подчеркнутые члены в сумме дают 0.

Таким образом, мы можем считать исходным уравнением уравнение, записанное в виде $x^3 + px + q = 0$. Сведем его к системе относительно неизвестных u и v, делая замену $x = u + v$:

$$\begin{aligned} &u^3 + 3u^2v + 3uv^2 + v^3 + p(u + v) + q = \\ &= u^3 + v^3 + 3uv(u + v) + p(u + v) + q = \\ &= u^3 + v^3 + (u + v)(3uv + p) + q. \end{aligned}$$

Если мы положим $3uv + p = 0$, то исходное уравнение можно заменить системой

$$\begin{cases} u^3 + v^3 = -q, \\ uv = -\dfrac{p}{3}. \end{cases}$$

Каждое решение этой системы $(u; v)$ дает решение $x = u + v$ исходного уравнения.

Система симметрична. Для ее решения возведем обе части второго уравнения в куб (при этом мы приобретаем «лишние» корни, аналогично тому, как мы их приобретали при возведении уравнения в квадрат).

$$\begin{cases} u^3 + v^3 = -q, \\ u^3v^3 = -\dfrac{p^3}{27}. \end{cases}$$

Числа u^3 и v^3 являются корнями квадратного уравнения $t^2 + qt - \frac{p^3}{27} = 0$. Запишем их с помощью квадратных радикалов, допуская запись $\sqrt{a}$, где $a < 0$: $t_{1,2} = -\frac{q}{2} \pm \sqrt{\frac{q^2}{4} + \frac{p^3}{27}}$.

(Мы при этом символом $\sqrt{a}$ фиксируем какое-либо «число», квадрат которого принимается равным a, записывая в виде $-\sqrt{a}$ другое число с тем же свойством.)

Чтобы найти u и v, надо извлечь кубические корни из чисел t_1 и t_2. Однако их нельзя извлекать как попало — надо будет проследить за тем, чтобы выполнялось равенство $uv = -\dfrac{p}{3}$.

Запишем ответ в символическом виде:

$$x = \sqrt[3]{-\frac{q}{2} + \sqrt{\frac{q^2}{4} + \frac{p^3}{27}}} + \sqrt[3]{-\frac{q}{2} - \sqrt{\frac{q^2}{4} + \frac{p^3}{27}}}.$$

Это и есть знаменитая формула Кардано.

Число $d = \dfrac{q^2}{4} + \dfrac{p^3}{27}$ играет роль дискриминанта. Если $d > 0$, то исходное уравнение (с вещественными p и q) имеет один вещественный корень (получающийся из формулы Кардано, если под $\sqrt[3]{a}$ понимать единственный существующий вещественный корень из вещественного числа a) и два сопряженных комплексных корня.

Если $d < 0$, то оказывается, что все три корня вещественны (в этом и был парадокс: вещественные числа были выражены через мнимые объекты).

Если $d = 0$, то число различных корней уменьшается: при $p \neq 0$ (а тогда и $q \neq 0$) получим два различных корня, а при $p = q = 0$ — один корень.

Основная теорема алгебры

Основным мотивом введения комплексных чисел была необходимость записывать корни кубических уравнений с вещественными коэффициентами. Возникают естественные вопросы — достаточно ли комплексных чисел, чтобы можно было найти корни любого алгебраического уравнения? А что если в качестве коэффициентов уравнения брать комплексные числа, не понадобятся ли новые числа для записи их корней? Ответом на оба эти вопроса является знаменитая теорема, доказанная немецким математиком К. Гауссом, которую часто называют основной теоремой алгебры.

Теорема Гаусса. Всякое алгебраическое уравнение степени $n \geq 1$ с комплексными коэффициентами имеет хотя бы один комплексный корень.

Доказательство этой теоремы использует свойства функций комплексного аргумента и по существу выходит за рамки алгебры.

Возьмем многочлен $f(x) = a_0x^n + a_1x^{n-1} + ... + a_{n-1}x + a_0$ с произвольными коэффициентами. Пусть x_1 — его комплексный корень, существующий по теореме Гаусса. Тогда многочлен $f(x)$ делится на $x - x_1$, т.е. $f(x) = (x - x_1)f_1(x)$, где степень многочлена $f_1(x)$ на единицу меньше степени $f(x)$. Если многочлен $f_1(x)$ не является константой (т.е. если его степень ≥ 1), то снова по теореме Гаусса мы найдем его корень x_2 и запишем равенства $f_1(x) = (x - x_2)f_2(x)$ и $f(x) = (x - x_1) \times \times (x - x_2)f_2(x)$. Этот процесс можно продолжать до тех пор, пока мы не придем к многочлену $f_n(x)$, являющемуся константой (равной коэффициенту a_0). В итоге получим разложение многочлена $f(x)$ на линейные множители:

$$f(x) = a_0(x - x_1)(x - x_2) \cdot ... \cdot (x - x_n).$$

Среди корней $x_1, x_2, ..., x_n$ могут оказаться равные, и в разложении многочлена $f(x)$ на линейные множители их можно объединить между собой.

Контрольные вопросы и задания

1. Дайте определение комплексного числа.
2. Что такое мнимая единица?
3. Как изобразить комплексное число в координатной плоскости?
4. Приведите примеры комплексных чисел и им противоположных.
5. Приведите примеры, как вычислить сумму и произведение двух комплексных чисел.
6. Перечислите законы сложения и умножения комплексных чисел.
7. Дайте определение числа, комплексно сопряженного с данным.
8. Приведите примеры на свойства 1—4 комплексного сопряжения.
9. Что такое модуль комплексного числа?
10. Приведите примеры на свойства модуля комплексных чисел.

§ 3. Понятие функции

Переменные

«Поворотным пунктом в математике была декартова переменная величина. Благодаря этому в математику вошли движение и тем самым диалектика и благодаря этому же стало немедленно необходимым дифференциальное и интегральное исчисление, которое тотчас и возникает и которое было в общем в целом завершено, а не изобретено Ньютоном и Лейбницем.» Эти слова принадлежат Фридриху Энгельсу. Они ярко характеризуют новый этап в развитии математики, который связан с именами великих ученых XVII в.: Декарта, Ньютона и Лейбница. На основе их работ сформировалось понятие функции, были разработаны методы исследования функций, которые в течение 300 лет остаются основным инструментом изучения окружающего мира с помощью математики.

Математика всегда была связана с вычислениями и формулами. Особенно много формул было получено при решении задач измерения — тысячелетия назад люди овладели формулами вычисления длин, площадей и объемов простейших фигур.

С помощью формул записываются соотношения между различными величинами.

Примеры

1. $V = \frac{4}{3}\pi R^3$ — объем шара.
2. $s = \frac{gt^2}{2}$ — путь при свободном падении.
3. $E = \frac{mv^2}{2}$ — кинетическая энергия.
4. $S = \sqrt{p(p-a)(p-b)(p-c)}$ — формула Герона для вычисления площади треугольника.
5. $x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$ — формула корней квадратного уравнения.

При выводе формул, при организации вычислений по формулам приходится делать преобразования выражений. Эта сторона математики читателю уже хорошо знакома. Она будет продолжена и в этом курсе, частью которого будет «алгебра», отражающая выполнение различных преобразований, т.е. операционную сторону математики.

Математический анализ, начала которого мы будем изучать в этом курсе, смотрит на формулу как на соотношение между меняющимися, переменными величинами. Как изменится точность вычисления объема шара, если точность измерения его радиуса изменить на одну сотую? Это типичный вопрос математического анализа. Ответ на него еще можно получить с помощью преобразований, так как формула объема шара не слишком сложна. Однако ответ на аналогичный вопрос для формулы Герона получить алгебраическими средствами трудно. Математический анализ создал методы, с помощью которых можно следить за характером изменения связанных между собой величин.

В первой из приведенных формул участвуют буквы V, R, π (V — это объем шара, R — его радиус, π — отношение длины окружности к диаметру). Величины V и R могут меняться, а π является постоянной. Приближенное значение константы π равно 3,14159 (с точностью до 0,00001). Во второй формуле встречаются буквы s, t, g (s — это длина пути, t — время движения, g — ускорение свободного падения). Величины s и t могут меняться, а величина g в этой формуле считается постоянной.

Переменная — это общий термин для обозначения различных меняющихся величин.

Например, рассматривая поведение газа в замкнутом объеме, можно измерить его температуру T, его объем V, оказываемое им давление p.

Наблюдая за свободно падающим телом, можно измерить длину пути s, пройденного телом за время t, его скорость v в момент времени t, его кинетическую энергию E в момент времени t и т.д.

В этих примерах участвуют различные переменные величины, или просто переменные.

Зависимости

Переменные, появляющиеся при описании какого-либо процесса, обычно связаны между собой. Одной из основных задач экспериментальных наук является изучение этих связей. Например, закон Клапейрона — Менделеева утверждает, что давление p, объем V и температура T идеального газа связаны соотношением $\frac{pV}{T} = \text{const}$.

Реальные процессы обычно связаны с большим количеством переменных и зависимостей между ними. В то же время можно отвлечься от частных деталей, сосредоточив свое внимание лишь на некоторых сторонах процесса, идеализировав условия, в которых он протекает. Тогда удается построить математическую модель процесса, состоящую в перечислении основных характеристик и тех связей, которые между ними имеются.

Например, физики, вводя понятие идеального газа, пренебрегают взаимодействием между молекулами газа и их размером и получают газовые законы в виде соотношений между переменными p, V и T.

При изучении падения материального тела можно пренебречь сопротивлением воздуха, изменением силы тяжести и др. и считать, что движение происходит по прямой с постоянным ускорением g. Тогда положение тела в любой момент времени t можно найти, зная его начальное положение и начальную скорость.

Какие значения могут принимать переменные? Во всех приведенных ранее примерах переменные величины были скалярными, т.е. их значения задаются числами. И в дальнейшем мы будем изучать в основном переменные, принимающие числовые значения.

Границы, в которых могут меняться числовые значения переменных, обычно определяются физическими условиями. Так, закон Клапейрона — Менделеева верен лишь при значениях p, V и T, лежащих в определенных промежутках, находимых экспериментально.

Изучение различных зависимостей между переменными является нашей главной задачей. Хотя величин очень много, но давно было замечено, что разные величины, появляющиеся при описании далеких друг от друга процессов, могут быть связаны между собой одной и той же зависимостью. Поэтому для изучения зависимостей полезно отвлечься от физического смысла рассматриваемых переменных.

Для нас в дальнейшем термин «переменная» будет означать букву, для которой указано множество принимаемых ею значений.

Например, мы будем говорить: рассмотрим переменную x, принимающую положительные значения.

В огромном море зависимостей между переменными можно выделить три типа простейших зависимостей, которые встречаются чаще всего, — это прямая и обратная пропорциональность и квадратичный закон.

Пусть x и y — две переменные.

1. Переменные x и y связаны прямо пропорциональной зависимостью, если их отношение постоянно.

С помощью формул эту зависимость можно записать так: $\frac{y}{x}=k$, или $y = kx$, где k — постоянное число, $k \neq 0$.

2. Переменные x и y связаны обратно пропорциональной зависимостью, если их произведение постоянно.

Запишем эту зависимость с помощью формул $xy = c$ или $y=\frac{c}{x}$, где c — постоянное число, $c \neq 0$.

3. Переменная y зависит от переменной x по квадратичному закону, если ее значения можно вычислить по формуле $y = ax^2$, где a — постоянное число, $a \neq 0$.

Заметим, что зависимость первых двух типов имеет симметричный характер — если переменная y прямо (или обратно) пропорциональна переменной x, то и, наоборот, переменная x прямо (или обратно) пропорциональна переменной y.

Например, если $y = \frac{3}{2}x$, то $x = \frac{2}{3}y$, или если $y = \frac{5}{x}$, то $x = \frac{5}{y}$.

Квадратичная зависимость не является симметричной, в нее переменные входят неравноправно. Если y зависит от x по квадратичному закону, то это неверно для зависимости x от y.

Графики простейших зависимостей приведены на рис. 12. Графиком прямо пропорциональной зависимости является прямая. График обратно пропорциональной зависимости называется гиперболой. График квадратичной зависимости — парабола.

Большинство известных зависимостей между физическими величинами принадлежит к одному из указанных нами простейших типов.

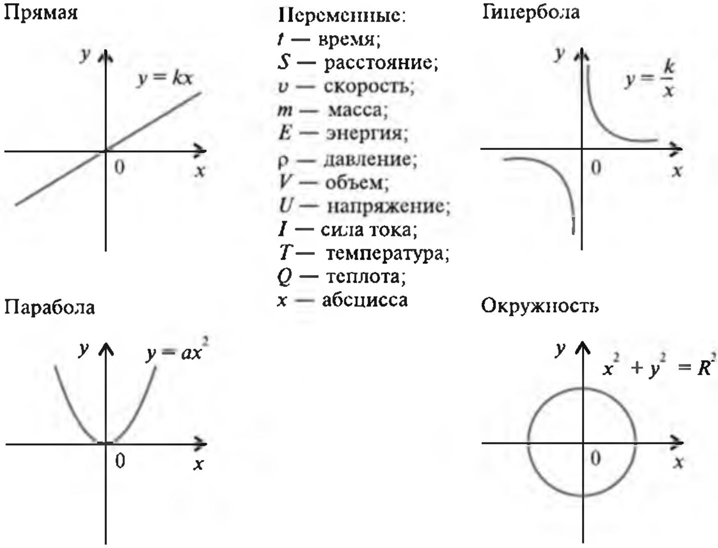


Рис. 12

Примеры

1. $U = RI$ — закон Ома, где U — напряжение, I — сила тока, R — сопротивление. Если одна из этих трех переменных постоянна, то

две другие связаны по закону Ома прямо или обратно пропорциональной зависимостью.

2. $pV = RT$ — закон Клапейрона — Менделеева, где p — давление, V — объем, T — температура, R — константа (газовая постоянная). При постоянной температуре давление и объем газа обратно пропорциональны друг другу (рис. 13).

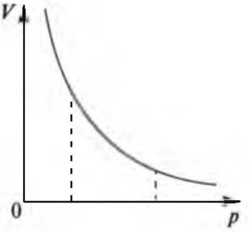


Рис. 13

3. $s = vt$ — закон равномерного движения, где s — путь, t — время, v — скорость.
4. $s = \frac{gt^2}{2}$ — закон свободного падения, где s — путь, t — время, g — константа.
5. $q = CU$ — уравнение конденсатора, где U — напряжение, q — заряд, C — емкость.
6. $E = \frac{mv^2}{2}$ — кинетическая энергия материальной точки, где v — скорость, m — масса.
7. $V = \frac{1}{3}\pi R^2 H$ — объем кругового конуса, где R — радиус основания, H — высота.
8. $m' = \frac{m}{v}$ — формула для нормы прибавочной стоимости, где m — прибавочный труд, v — необходимый труд.

Определение функции

Пусть даны две переменные x и y. Говорят, что переменная y является функцией от переменной x, если задана такая зависимость между этими переменными, которая позволяет для каждого значения x однозначно определить значение переменной y.

Примеры функций

1. $y = kx + b$.
2. $y = |x|$.
3. $y = x^2$.
4. $y = \frac{1}{x}$, $x > 0$.

В каждом из этих примеров указана формула, позволяющая для каждого значения переменной x однозначно вычислить значение переменной y.

Для того чтобы задать функцию, нужно указать:

1) множество всех возможных значений переменной x. Это множество, которое мы будем обозначать D, называют областью определения функции;
2) правило, по которому каждому числу x из множества D сопоставляется число y, определяемое числом x. Это число y называется значением функции в точке x.

Функция обычно обозначается одной буквой, например f Значение функции f в точке x обозначается $f(x)$.

Итак, если задана функция f, то задано множество чисел $D = D(f)$ и каждому числу $x \in D$ сопоставлено число $y = f(x)$.

Переменную x называют аргументом. D — это область значений аргумента.

Пусть задана функция f с областью определения D. Рассмотрим плоскость с декартовой системой координат. По оси абсцисс будем откладывать значение аргумента, а по оси ординат — значение функции. Для каждого числа $x \in D$ вычислим $y = f(x)$ и построим точку M с координатами $(x, f(x))$ (рис. 14). Множество всех таких точек образует кривую, называемую графиком функции f в заданной системе координат.

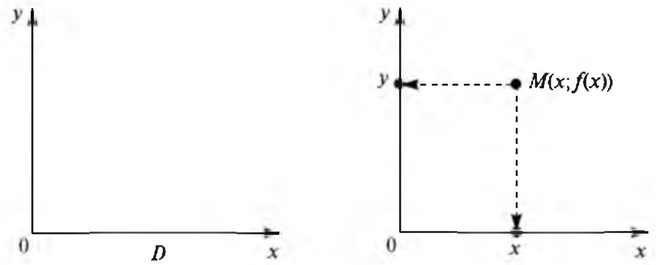


Рис. 14

Итак, графиком функции f называется множество точек плоскости с координатами $(x, f(x))$, где x пробегает область определения функции f.

На рисунках 15 и 16 изображены графики функций, которые были приведены здесь в качестве примера.

Изученные ранее простейшие зависимости определяют три важнейшие функции:

1) $y = kx$; 2) $y = \dfrac{c}{x}$; 3) $y = ax^2$.

Эти функции являются представителями функций из трех классов, которые мы будем в дальнейшем рассматривать: линейных, дробно-линейных и квадратичных.

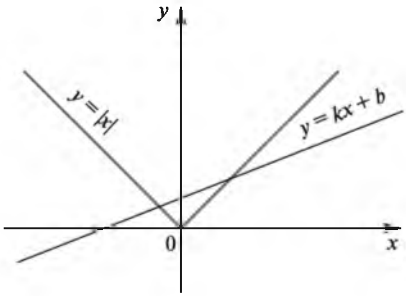


Рис. 15

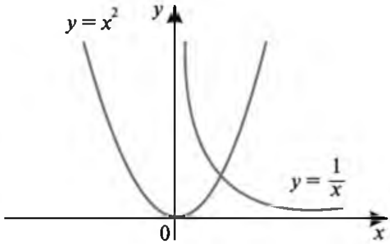


Рис. 16

Способы задания функции

Для того чтобы определить переменную y как функцию от переменной x, нужно задать множество значений аргумента x и указать правило вычисления значений y в зависимости от x. Сначала обсудим, как задается правило вычисления значений. Во всех приведенных ранее примерах правило вычисления задавалось явной формулой, содержащей определенные операции.

Обучаясь математике, мы ознакомились с различными действиями, операциями над числами. Например, используя только сложение и умножение, можно из числа x получить новые числа, скажем, $3x$, $3x + 5$, $x^3 + 3x + 5$ и др. Такие выражения, многочлены, могут служить для построения довольно богатого запаса функций.

Использование деления расширяет этот запас. Теперь можно образовать выражения вида $\frac{1}{x}$, $\frac{x+1}{x}$, $\frac{2x}{x^2-1}$ и др. Построенные таким образом функции называют *рациональными*.

Операция деления отличается от сложения и умножения тем, что она не всегда определена — в знаменателе дроби нельзя ставить нуль. Поэтому, например, в выражение $\frac{2x}{x^2-1}$ можно подставить любые числа, кроме $x = 1$ и $x = -1$, при которых знаменатель равен нулю.

Появление новых операций и введение специальных знаков для их обозначения приводит к дальнейшему обогащению наших возможностей — извлечение корня, нахождение синуса, переход к модулю числа и т.п.

Например, пусть $f(x)$ равно -1, если $x < 0$, равно 0, если $x = 0$, и равно $+1$, если $x > 0$. Этими словами мы описали правило вычисления, применимое к любому числу. Обозначим число $f(x)$, найденное по этому правилу, через $\operatorname{sgn} x$ (от лат. *signum* — знак). Теперь с помощью символа для обозначения новой операции можно строить новые формулы, например $g(x) = \frac{1}{x} + \operatorname{sgn}(x-1)$.

Если функция задана формулой и не указаны никакие ограничения, то ее областью определения считается множество всех значений аргумента, при которых выполнимы все операции, участвующие в формуле. Это множество называют *естественной областью* определения данной функции.

Так, естественной областью определения функции $y=\sqrt{1-x^2}$ является множество чисел x, для которых $1 - x^2 \geq 0$, т.е. промежуток $[-1, 1]$.

Еще раз обратим внимание на то, что две важные операции — деление и извлечение корня четной степени — выполнимы не всегда (нельзя разделить на нуль, нельзя извлечь корень четной степени из отрицательного числа). Это ограничение надо помнить и учитывать при нахождении области определения функции, в задании которой участвуют указанные операции.

Значения функции $y=\sqrt{x^2+1}$ вычисляются путем последовательного выполнения операций: возведения в квадрат, прибавления единицы, извлечения квадратного корня. Можно сказать, что $y=\sqrt{x^2+1}$ является «сложной функцией», составленной из более простых: $u = x^2$; $v = u + 1$; $y=\sqrt{v}$.

Итак, правила вычисления значений функции могут задаваться формулами, полученными с помощью действий над числами.

Другой важный способ задания значений функции — табличный. В таблице можно непосредственно указать значения функции, однако лишь для конечного набора значений аргумента.

Вычисление значений функции может быть запрограммировано в калькуляторе. Вычислительное устройство может служить способом задания новой функции. Современные вычислительные машины снабжены клавишами, позволяющими немедленно вычислить значения многих функций.

Наконец, часто функцию задают с помощью графика. Этот способ очень удобен — он позволяет наглядно представить свойства функции.

Примеры

На рисунке 17 изображены вольт-амперные характеристики некоторых электрических элементов, т.е. графически заданные зависимости напряжения от силы тока. Они получены не по готовой формуле, а экспериментально.

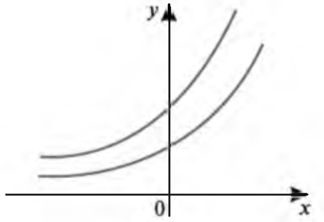


Рис. 17

Рис. 18

На рисунке 18 изображена кардиограмма работы человеческого сердца. Ее можно считать графиком изменения электрического потенциала U на волокнах сердечной мышцы во время сердечного цикла.

Контрольные вопросы и задания

1. Обратите внимание на следующие ключевые слова и обозначения, появившиеся в тексте: функция, область определения функции, график функции $y = f(x)$. Приведите примеры их использования.
2. Что нужно указать для задания функции?
3. Что такое график функции?
4. Какие вы знаете способы задания функции?
5. Что понимают под областью определения функции, заданной формулой?
6. Приведите пример линейной функции.
7. Приведите пример дробно-линейной функции.
8. Приведите пример квадратичной функции.
9. Приведите пример функции, естественная область определения которой меньше, чем вся числовая ось x.
10. Приведите пример графического задания функции.

§ 4. Исследование функции

Чтение графика

Рассмотрим функцию f, график которой показан на рис. 19. О каких свойствах функции f можно сказать, глядя на график?

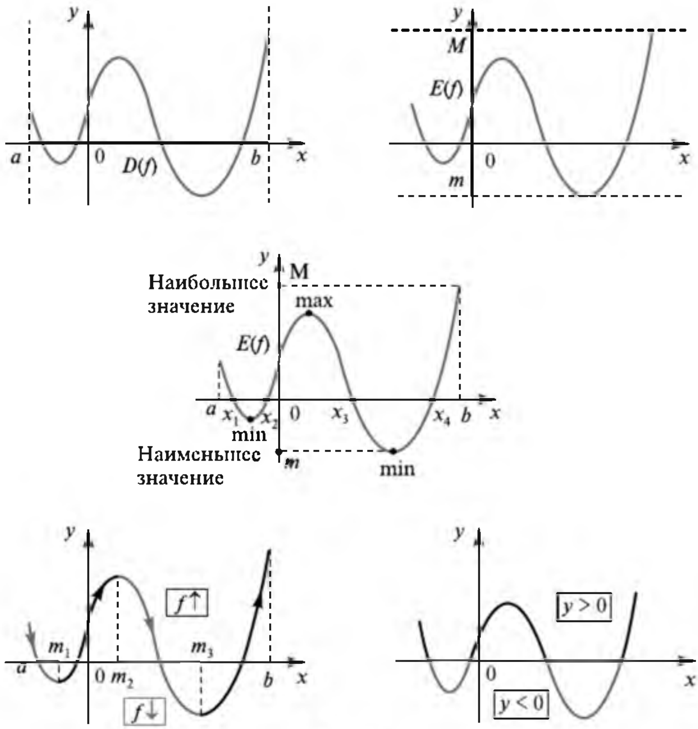


Рис. 19

1. Спроектируем точки графика на ось x. Мы получим отрезок $[a; b]$. Он является областью определения функции $D(f) = [a; b]$. Каждая прямая, параллельная оси y и проходящая через точку этого отрезка, пересекает график в одной точке. Вертикальные прямые, проходящие через точки x вне отрезка $[a, b]$, график не пересекают.

2. Рассмотрим точки пересечения графика с осью x: x_1, x_2, x_3, x_4. В этих точках функция обращается в нуль. Числа x_1, x_2, x_3, x_4 являются корнями уравнения $f(x) = 0$. Итак, множество чисел $\{x_1, x_2, x_3, x_4\}$ представляет собой множество решений уравнения $f(x) = 0$. Указанные числа называются *нулями функции* (или *корнями функции*).

3. *Нули* функции f разбивают отрезок $[a, b]$ на промежутки, в каждом из которых функция f сохраняет постоянный знак. Функция положительна на промежутках (a, x_1), (x_2, x_3), (x_4, b) и отрицательна на промежутках (x_1, x_2), (x_3, x_4).

Объединение промежутков $(a, x_1) \cup (x_2, x_3) \cup (x_4, b)$ представляет собой решение неравенства $f(x) > 0$.

4. Из рисунка 19 видно, какую роль для определения формы графика играют «горбы» функции f Они соответствуют точкам m_1, m_2, m_3. Выясним, какие свойства функции характеризуют «горбы» и «впадины».

Точка m_1 замечательна тем, что если мы рассмотрим ход изменения функции f вблизи нее, то значение f в этой точке будет наименьшим. Отметим еще раз: $f(m_1)$ не является самым маленьким значением функции f (на чертеже легко найти точки, в которых значение f меньше, чем $f(m_1)$). Это наименьшее значение среди значений в точках, близких к m_1. Точно так же в точке m_2 функция принимает значение, наибольшее среди значений в точках, близких к m_2.

Точку m_1 называют точкой *локального минимума*, а точку m_2 — точкой *локального максимума*. Слово «локальный» означает «местный» и подчеркивает, что $f(m_1)$ — наименьшее значение вблизи точки m_1.

Итак, в точках m_1 и m_3 функция f имеет локальный минимум, а в точке m_2 — локальный максимум.

5. Наибольшее значение функция f принимает в точке b, а наименьшее — в точке m_3. На этот раз речь идет о самом большом и самом малом значениях функции на всей области определения.

Латинские слова «максимум» и «минимум» соответствуют русским словам «наибольшее значение» и «наименьшее значение». Иногда латинские слова оставляют для обозначения «горбов» (и даже опускают слово «локальный») и просто говорят: точки максимумов и минимумов функции, а русские слова — для обозначения общего, глобального, абсолютного максимума и минимума. Часто используют еще одно латинское слово — «экстремум», которое объединяет оба понятия: максимум и минимум.

6. Что можно сказать о поведении функции f на промежутках между точками экстремумов?

Из графика ясно, что на отрезке $[a, m_1]$ функция f убывает, затем на отрезке $[m_1, m_2]$ она возрастает, на отрезке $[m_2, m_3]$ функция f убывает и, наконец, на отрезке $[m_3, b]$ снова возрастает.

7. Спроектируем точки графика на ось y. Мы получим отрезок $[m, M]$, являющийся областью значений функции f. Он состоит из всех точек y, являющихся значениями функции f при каком-либо (не обязательно одном) значении x.

Свойства функции

1. ***Область определения*** — множество значений аргумента, при которых задана функция.

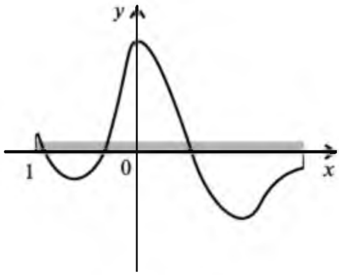


Рис. 20

Геометрически — это проекция графика функции на ось абсцисс (рис. 20).

Если функция задана формулой, то имеется в виду ее *естественная область определения,* т.е. множество чисел, к которым применима данная формула. Ответ можно привести, записывая отрезок числовой оси с помощью неравенств или указывая «исключительные» значения аргумента (значения, которые не входят в область определения функции).

Например: 1) $f(x) = \dfrac{x}{x^2+1}$; $D(f) = \mathbb{R}$; 2) $f(x) = \dfrac{1}{x}$; $D(f)$: $x \neq 0$, или $D(f)$: $(-\infty; 0) \cup (0; +\infty)$; 3) $f(x) = \sqrt{x}$; $D(f)$: $[0; +\infty)$, или $x \geq 0$.

2. ***Нули функции*** — это точки, в которых функция обращается в нуль, или, иначе говоря, решения уравнения $f(x) = 0$.

Геометрически — это точки пересечения графика функции с осью абсцисс (рис. 21).

3. ***Знакопостоянство*** — это промежутки, на которых функция положительна или отрицательна, или, иначе, решения неравенств $f(x) > 0$ и $f(x) < 0$.

Геометрически — это участки оси абсцисс, соответствующие точкам графика, лежащим выше (ниже) этой оси (рис. 22).

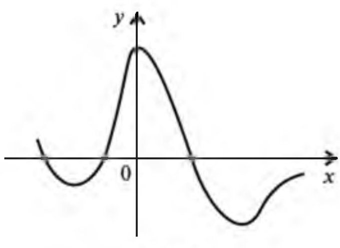


Рис. 21

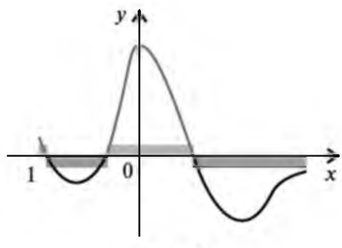


Рис. 22

Промежутки знакопостоянства функции, рассмотренной ранее, определялись ее корнями (см. рис. 19). Однако из этого правила бывают исключения. Такие графики показаны на рис. 23, *а—в*.

В случаях *а* и *б* на рис. 23 функция изменила свой знак, но в нуль не обратилась. Это бывает тогда, когда график функции разорван. Особенно характерным является случай *а*, здесь в точке $x = 0$ функция просто не определена. Случай *в* показывает, что функция может обратиться в нуль, но знака при этом не изменяет.

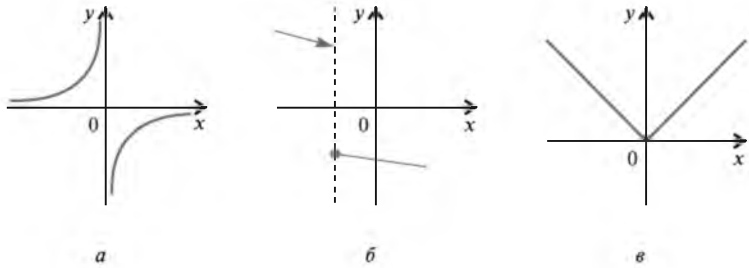


Рис. 23

4. *Точки экстремума* — точки, лежащие внутри области определения, в которых функция принимает самое большое (максимум) или самое малое (минимум) значение по сравнению со значениями в близких точках.

Геометрически — это точки оси абсцисс, в которых график функции имеет «горб» (рис. 24).

Рис. 24

В простейших случаях точки экстремума находятся непосредственно. Нахождение точек экстремума более сложных функций делается с помощью производной.

5. ***Монотонность.*** Под монотонностью функции понимают ее возрастание или убывание.

Уточним понятие монотонности функции. При грамотном использовании терминов «возрастание» и «убывание» функции надо обязательно указывать, на каком участке изменения аргумента рассматривается монотонность функции. Мы будем использовать понятие функции, монотонной на промежутке.

Рассмотрим функцию f и некоторый промежуток, целиком входящий в область определения функции f. Обозначим выбранный промежуток одной буквой, например A.

Функция f называется возрастающей (убывающей) на промежутке A, если для любых двух чисел x_1 и x_2 из этого промежутка, таких что $x_1 < x_2$, выполняется условие $f(x_1) < f(x_2)$ (соответственно, $f(x_1) > f(x_2)$).

Символически определение возрастания функции может быть записано так: $x_1 < x_2 \Rightarrow f(x_1) < f(x_2)$. Символ $\Rightarrow$ мы будем использовать для обозначения условного утверждения: если.., то...

Примеры

1. Функция $y = 2x$ является возрастающей на всей числовой оси (и, разумеется, на каждом ее промежутке).
2. Функция $y = 3x^2$ является убывающей на промежутке $(-\infty, 0]$ и возрастающей на промежутке $[0, +\infty)$.

Функция $y = \frac{1}{x}$ является убывающей как на промежутке $(-\infty, 0)$, так и на промежутке $(0, +\infty)$.

Подчеркнем еще раз, что мы будем говорить о монотонности функции только на промежутке, целиком входящем в ее область определения. Поэтому нельзя сказать в целом о функции $y = \frac{1}{x}$, что она монотонна, хотя она убывает на каждом промежутке, где она определена.

Все функции, с которыми мы будем в дальнейшем встречаться, склеены из функций, монотонных на промежутках. Иначе говоря, мы будем изучать такие функции, область определения которых можно разбить на промежутки, на каждом из которых функция монотонна.

Разберемся в том, что может происходить при склеивании функций, монотонных на двух соприкасающихся промежутках A и B.

Все интересующие нас случаи представлены на рис. 25 и 26. Прежде всего, наглядно видна разница между двумя группами графиков.

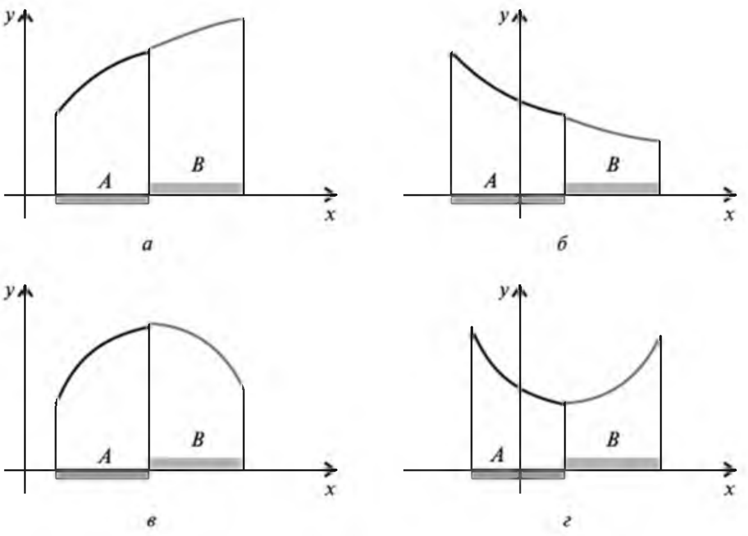


Рис. 25

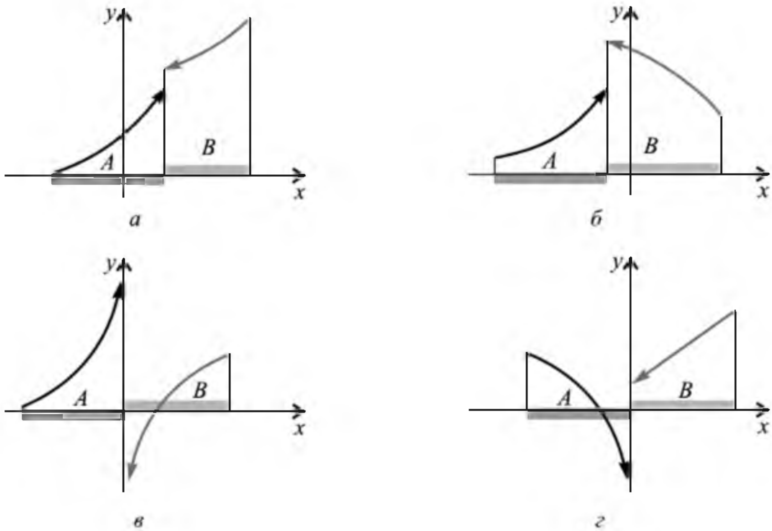


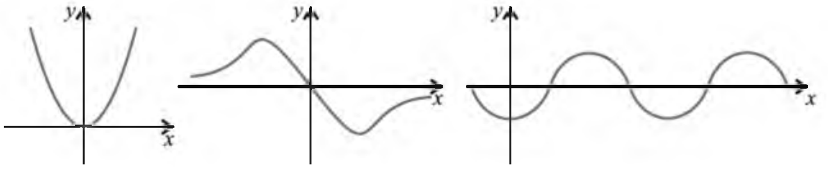
Рис. 26

Все функции, изображенные на рис. 25, определены в точке стыковки промежутков A и B. Кроме того, графики этих функций не имеют разрывов в отличие от всех графиков, изображенных на рис. 26.

Ситуация, изображенная на рис. 25, ясна. В случаях *а* и *б* промежутки A и B надо объединить. Функция останется монотонной в объединенном промежутке. В случае *в* возрастание функции на промежутке A сменяется убыванием на промежутке B. В точке стыковки функция имеет максимум. Наоборот, в случае *г* в точке стыковки функция имеет минимум.

Во всех случаях, изображенных на рис. 26, функция имеет разрыв в точке стыковки промежутков монотонности. Такую точку стыковки мы будем называть ***особой точкой***. Поведение функции вблизи особой точки может быть различным, как это видно из рис. 26. При построении графиков функций особые точки придется исследовать отдельно.

6. ***Симметрия графика.*** Часто график функции является симметричным. На рисунке 27 представлены различные виды симметрии графика. В случае *а* график обладает осевой симметрией (относительно оси y); в случае *б* график симметричен относительно начала координат; в случае *в* график периодически повторяется. Первое из этих свойств называется ***четностью***, второе — ***нечетностью***, третье — ***периодичностью***.



а *б* *в*

Рис. 27

Функция $y = f(x)$ называется *четной*, если при всех значениях аргумента $f(-x) = f(x)$.

При этом имеется в виду, что если x входит в область определения, то и $-x$ также входит в нее, т.е. что область определения функции f симметрична относительно начала координат.

Так как точки $(x; y)$ и $(-x; y)$ одновременно принадлежат или не принадлежат графику четной функции, то этот график симметричен относительно оси y.

Аналогично, функция $y = f(x)$ с областью определения, симметричной относительно начала координат, называется *нечетной*, если при всех значениях аргумента $f(-x) = -f(x)$.

Так как точки $(x; y)$ и $(-x; -y)$ одновременно принадлежат или не принадлежат графику нечетной функции, то график симметричен относительно начала координат.

Функция $y = f(x)$ называется *периодической*, если существует такое число $T > 0$, что при всех x выполняются равенства $f(x) = f(x + T) = f(x - T)$.

Разумеется, при этом имеется в виду, что область определения функции вместе с точкой x содержит и точки $x \pm T$.

Если число $T > 0$ является периодом функции f, то и числа $2T$, $3T$ и др., т.е. числа nT при любом натуральном n, являются периодами этой функции. Как правило, у функции сущеcтвет наименьший положительный период. Если T — наименьший положительный период функции f, то все ее другие периоды имеют вид nT, где $n \in N$.

Действительно, если $U > T$ — период, больший T, то мы запишем равенство $U = nT + V$, где $n \in \mathbb{N}$ и $0 < V < T$, V можно найти, откладывая на отрезке длины U отрезок T столько, сколько это возможно. $V = U - nT$ также будет положительным периодом. Неравенство $V < T$ противоречит выбору T в качестве наименьшего положительного периода.

Тригонометрические функции синус, косинус, тангенс и котангенс являются периодическими (о них далее в главе 3). Приведем еще один важный пример периодической функции. Рассмотрим функцию $y = x - [x]$, где через $[x]$ обозначена целая часть числа. Эта функция является периодической с наименьшим положительным периодом, равным 1:

$$f(x+1) = x + 1 - [x+1] = x + 1 - [x] - 1 = x - [x] = f(x).$$

График этой функции изображен на рис. 28.

7. ***Ограниченность функции.*** Функция $y = f(x)$ с областью определения $D(f)$ называется *ограниченной*, если существует такая константа M, что для всех $x \in D(f)$ выполняется неравенство $|f(x)| \le M$.

Можно рассчитывать одностороннюю ограниченность, требуя существования констант M и N, таких, чтобы выполнялись неравенства:

$f(x) \le M$ — ограниченность сверху; $f(x) \ge M$ — ограниченность снизу.

Легко проверить, что одновременная ограниченность сверху и снизу равносильна общему понятию ограниченности.

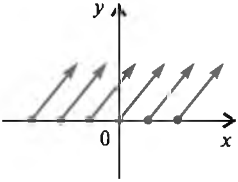


Рис. 28

Примеры

1. Функция $y = f(x)$, где $f(x) = \dfrac{1}{1+x^2}$ ограничена, так как $1 + x^2 \ge 1$ и $\dfrac{1}{1+x^2} \le 1$ (рис. 29). Модуль писать не нужно, так как $\dfrac{1}{1+x^2} > 0$.

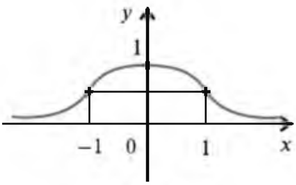


Рис. 29

2. Функция $y = 1 + x^2$ не ограничена, так как может принимать сколь угодно большие значения, но она является ограниченной снизу, так как $1 + x^2 \ge 1$ (рис. 30).

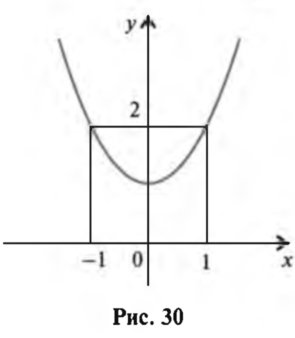


Рис. 30

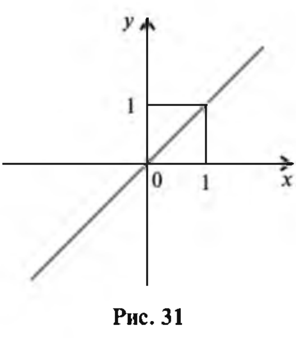


Рис. 31

Аналогично функция $y = 1 - x^2$ ограничена сверху: $1 - x^2 \leq 1$, но не ограничена снизу.

3. Функция $y = x$ не ограничена ни сверху, ни снизу (рис. 31).

8. ***Непрерывность.*** Ранее были (см. рис. 26) изображены графики функций, которые мы назвали разрывными. Попробуем более точно определить, что такое непрерывная функция.

Если в данной точке у функции разрыв, то это означает, что при маленьком сдвиге значения аргумента значение функции совершит скачок. Наоборот, функция непрерывна в данной точке, если при маленьком изменении аргумента мало будут меняться ее значения.

Какие разрывы могут встретиться у изучаемых нами функций? Разобьем область определения функции на промежутки монотонности и отдельно рассмотрим разрыв внутри промежутка монотонности и в точке стыковки.

Пусть функция f монотонна (например, возрастает) на конечном промежутке $[a, b]$. Пусть $f(a) = c$ и $f(b) = d$. Непрерывность функции f означает, что при изменении аргумента от a до b она принимает без пропусков все промежуточные значения от c до d. Разрывность монотонной функции означает наличие скачков, пробелов, пропусков среди ее значений (см. рис. 26, *а*).

Разрывы в точках стыковки промежутков монотонности могут быть устроены сложнее. Чаще всего у нас будут встречаться «бесконечные» разрывы, аналогичные разрыву в точке $x = 0$ у функции $y = \dfrac{1}{x}$.

Пусть функция f монотонна и непрерывна на отрезке $[a, b]$. Как мы уже сказали, это означает, что она принимает все промежуточные значения. Отсюда следует, что если на концах отрезка функция f принимает значения разных знаков, то в некоторой точке она обращается в нуль (рис. 32). Это свойство непрерывной функции часто используется при решении уравнений с помощью графика.

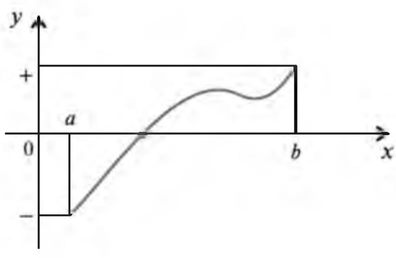


Рис. 32

Схема исследования функции

Определенные нами возможные свойства функции полезно собрать в одну схему, по которой полезно проводить исследование определенных функций. Заметим, что после изучения производной мы сможем проводить полное исследование и построение графиков более сложных функций, чем мы это делаем в настоящее время.

Контрольные вопросы и задания

1. Проверьте, понимаете ли вы смысл следующих ключевых слов: экстремум, максимум, минимум, монотонность, четность, нечетность, непрерывность.
2. Как с помощью графика найти область определения функции?
3. Как графически изображаются корни (нули) функции?
4. Как с помощью графика найти промежутки, на которых функция сохраняет постоянный знак?
5. Что такое точка локального максимума (минимума)?
6. Чем отличается наибольшее значение функции от ее локального максимума?
7. Перечислите, какие свойства функции включаются в схему ее исследования.
8. Дайте определение четной и нечетной функции. Какой симметрией обладают графики этих функций?
9. Дайте определение периодической функции. Как выглядит график такой функции?
10. Какая функция называется ограниченной?

§ 5. Операции над функциями и их графиками

Арифметические операции

Над функциями можно производить арифметические операции — сложение, умножение, деление. Пусть даны две функции f и g *с одной и той же областью определения* D.

Суммой функций f и g называется функция $h = f + g$, с той же областью определения, значения которой в точке x получаются сложением значений функций f и g в этой точке:

$$h(x) = f(x) + g(x).$$

Аналогично определяется умножение функций.

При делении функций возникают проблемы — надо требовать, чтобы функция, попадающая в знаменатель, не обращалась в нуль.

Итак, при определении арифметических операций над функциями у нас возникло два ограничения — у них должна быть одинаковая область определения, а при делении на функцию еще нужно дополнительно требовать, чтобы она не обращалась в нуль.

Что же делать, чтобы избавиться от этих ограничений?

Простейший способ — уменьшить, сузить область определения функций. Так, чтобы сложить функции $y = x$ и $y = \sqrt{x}$, у первой из которых область определения — $\mathbb{R}$, а у второй — промежуток $x \geq 0$, надо сначала область определения первой функции ограничить и считать ее такой же, как у второй функции. В итоге функция $y = x + \sqrt{x}$ считается заданной при $x \geq 0$.

Аналогично при делении функции $y = x^2$ на функцию $y = x - 1$ мы должны исключить из области определения первой функции точку $x = 1$, в которой вторая функция обращается в нуль. В итоге функция $y = \dfrac{x^2}{x-1}$ считается определенной при всех $x \neq 1$.

С помощью арифметических операций над функциями из постоянных функций и простейшей функции $y = x$ можно построить широкий класс функций, которые задаются формулами вида $f(x) = \dfrac{P(x)}{Q(x)}$, где $P(x)$ и $Q(x)$ — многочлены с буквой x. Такие функции называются ***рациональными***. Они определены во всех точках, при которых $Q(x) \neq 0$. При $Q(x) \equiv 1$ рациональные функции становятся многочленами.

Композиция функций

Чаще всего встречаются функции, заданные формулой. Формула строится путем последовательного выполнения различных известных операций над аргументом и постоянными числами. Эта процедура может быть уточнена с помощью понятия сложной функции.

Пусть даны две функции: $y = g(x)$ и $z = f(y)$. Сложной функцией (или композицией функции f и g) называется функция $z = h(x)$, значения которой вычисляются по правилу $h(x) = f(g(x))$, т.е. сначала вычисляется $g(x)$, при этом получается некоторое число y, а затем вычисляется значение f в точке y.

Со сложными функциями мы, разумеется, встречались и раньше. Так, функцию $z = \sqrt{1-x^2}$ можно рассматривать как композицию функций $y = 1 - x^2$ и $z = \sqrt{y}$.

Для организации вычислений по сложно устроенной формуле надо суметь разобраться в последовательности производимых операций, т.е. представить сложную функцию как композицию более простых. При этом встречается несколько трудностей. Первая из них связана с обозначениями. В приведенном ранее определении использованы три переменных: x, y и z. Ясно, что если y есть функция от x, а z есть функция от y, то z можно рассматривать как функцию от x. Однако часто приходится составлять сложную функцию из двух функций, обозначенных одной и той же переменной. Скажем, как составить композицию функций $y = x^2$ и $y = \dfrac{x}{x+1}$, или в общем виде $y = f(x)$ и $y = g(x)$? Для этого надо уточнить порядок применения данных функций, например, сначала g, а потом f или, наоборот, сначала f, а затем g. Для приведенного примера мы получим две различные функции: $y = \left(\dfrac{x}{x+1}\right)^2$ и $y = \dfrac{x^2}{x^2+1}$, а в общем виде $y = f(g(x))$ и $y = g(f(x))$.

Для записи сложной функции употребляется значок $\circ$, например запись $h = f \circ g$ означает, что функция h получена как композиция функций g и f (сначала применяется g, а затем f, т.е. $(f \circ g) = f(g(x))$. Как показывает приведенный пример, операция образования сложной функции (или операция композиции функций) не обладает переместительным свойством: $f \circ g \neq g \circ f$.

Другая трудность в обращении со сложными функциями состоит в том, что при этом необходимо следить за областями определения. Чтобы можно было вычислить сложную функцию $h = f(g(x))$, надо, чтобы число $g(x)$, т.е. значение функции g, попадало в область определения функции f. Так, вычисляя значения функции $y = \sqrt{1-x^2}$, мы должны брать только те числа x, для которых $1 - x^2 \geq 0$, т.е. те, для которых число $1 - x^2$ попадает в область определения функции $y = \sqrt{x}$.

Обратная функция

Формула для вычисления площади круга по его радиусу: $S = \pi R^2$. Эта формула задает площадь круга S как функцию его радиуса R, т.е. для каждого (положительного) числа по этой формуле вычисляется площадь круга S. Представим себе, что надо решить обратную задачу: по данной площади круга S вычислить его радиус. Для этого выразим R через S так: $R = \sqrt{\frac{S}{\pi}}$. Новая формула задает радиус круга R как функцию его площади S. Полученные две функции $S = S(R)$ и $R = R(S)$ являются примерами взаимно обратных функций.

Примеры

1. $y = 2x + 5$ и $x = \frac{y}{2} - \frac{5}{2}$.
2. $y = x^3, x \geq 0$ и $x = \sqrt[3]{y}$.
3. $y = \frac{x}{x-2}$ и $x = \frac{2y}{y-1}$.

В каждом из указанных примеров соответствие между переменными величинами, задаваемое взаимно обратными функциями, одно и то же. В самом деле, зависимость между радиусом и площадью круга остается одной и той же: записывается ли она в виде $S = \pi R^2$ или же в виде $R = \sqrt{\frac{S}{\pi}}$. Точно так же функции $y = 2x + 5$ и $x = \frac{y}{2} - \frac{5}{2}$ выражают одну и ту же зависимость между переменными x и y. Аналогично функции третьего примера происходят от одной и той же зависимости, которую можно записать так: $xy = x + 2y$.

Две функции f и g называются взаимно обратными, если формулы $y = f(x)$ и $x = g(y)$ выражают одну и ту же зависимость между переменными x и y, т.е. если равенство $y = f(x)$ верно тогда и только тогда, когда верно равенство $x = g(y)$:

$$y = g(x) \Leftrightarrow x = g(y).$$

Если две функции f и g взаимно обратны, то g называют обратной функцией для f и, наоборот, f — обратной для g.

Как ответить на вопрос: что такое обратная функция для функции f? Это можно сделать следующим образом: обратная функция для функции f — это такая функция g, что f и g образуют пару взаимно обратных функций.

Как мы уже отметили, зависимость $y = 2x + 5$ и $x = \frac{y}{2} - \frac{5}{2}$ между переменными x и y одна и та же. Эту же зависимость можно записать и так: $2x - y + 5 = 0$. Из последней формулы можно выразить y как функцию от x, а можно и, наоборот, выразить x как функцию от y. Эти две функции будут взаимно обратны.

Таким образом, исходным понятием является понятие зависимости. Если есть некоторая зависимость между переменными x и y, которая позволяет выразить y как функцию от x и x как функцию от y, то эти две функции и являются взаимно обратными.

Графики и свойства взаимно обратных функций

При построении графиков взаимно обратных функций необходимо внимательно следить за обозначениями переменных. Рассмотрим функцию f. Аргумент этой функции и ее значения можно обозначить произвольными буквами. Так, в формулах $y=x^2$, $S=t^2$, $N=n^2$ переменные обозначены различными буквами, однако каждая из этих формул определяет одну и ту же квадратичную функцию.

Определение взаимно обратных функций сделано нами на языке зависимостей. Чтобы определить, являются ли эти две функции f и g (заметьте, здесь пока нет обозначений для переменных) взаимно обратными, надо взять две переменные, например x и y, составить две формулы $y=f(x)$ и $x=g(y)$ и затем определить, задают эти две формулы одну и ту же зависимость между переменными x и y или нет.

Различия в обозначениях переменных сказываются при построении графиков функций. Пусть у нас есть две переменные x и y, значения которых откладываются на выбранных координатных осях, которые мы обозначаем этими же буквами x и y. Рассмотрим зависимости между x и y: $y=2x+5$; $x=\frac{y}{2}-\frac{5}{2}$; $2x-y+5=0$.

Ясно, что это разные записи одной и той же зависимости между переменными x и y:

$$y=2x+5 \Leftrightarrow x=\frac{y}{2}-\frac{5}{2} \Leftrightarrow 2x-y+5=0.$$

Поэтому график каждой из этих зависимостей один и тот же — он состоит из всех точек $P(x, y)$, координаты которых связаны соотношением $y = 2x + 5$ справедливым тогда и только тогда, когда $x = \frac{y}{2}-\frac{5}{2}$ или $2x - y + 5 = 0$ (рис. 33).

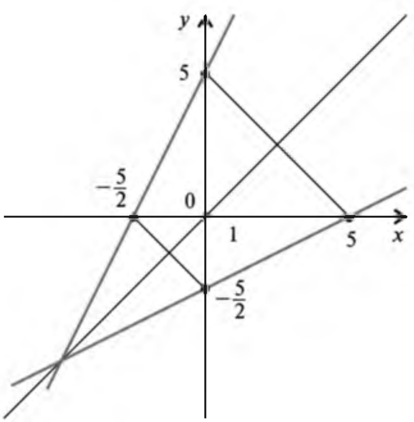


Рис. 33

Посмотрим на зависимость $x = \frac{y}{2} - \frac{5}{2}$. Она выражает x как некоторую функцию от y. Аргумент этой функции обозначен буквой y, а значения — буквой x. Поменяем местами x и y, т.е. рассмотрим зависимость $y = \frac{x}{2} - \frac{5}{2}$. Функция осталась одной и той же, но теперь ее аргумент обозначен, как обычно, буквой x, а значения — буквой y. Построим график функции $y = \frac{x}{2} - \frac{5}{2}$ (например, по двум точкам, см. рис. 33). Мы видим, что этот график отличен от графика зависимости $x = \frac{y}{2} - \frac{5}{2}$.

Как связаны между собой графики функций $y = 2x + 5$ и $y = \frac{x}{2} - \frac{5}{2}$? Возьмем какую-либо точку на графике первой функции, например $P\,(0;\,5)$. Поменяем местами координаты, т.е. рассмотрим точку $Q\,(5;\ 0)$. Эта точка лежит на графике второй функции $0 = \frac{5}{2} - \frac{5}{2}$. Точки $P\,(0;\,5)$ и $Q(5;\,0)$ симметричны друг другу относительно биссектрисы угла xOy, т.е. прямой $y = x$. Из рисунка 34 видно, что при любых a и b точки $P(a;\,b)$ и $Q(b;\,a)$ симметричны друг другу относительно прямой $y = x$.

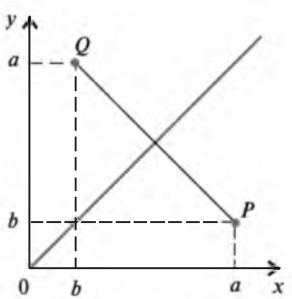


Рис. 34

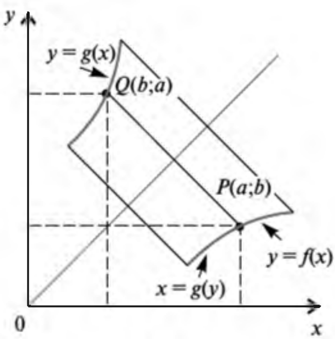


Рис. 35

Теорема. Пусть f и g — взаимно обратные функции. Графики функций $y = f(x)$ и $y = g(x)$ симметричны друг другу относительно биссектрисы угла xOy.

Доказательство. По определению взаимно обратных функций, формулы $y = f(x)$ и $x = g(y)$ выражают одну и ту же зависимость между переменными x и y, а значит, эта зависимость изображается одним и тем же графиком — некоторой кривой C (рис. 35). Кривая C является графиком функции $y = f(x)$. Возьмем произвольную точку $P(a;\ b) \in C$. Это означает, что $b = f(a)$ и одновременно $a = g(b)$. Построим точку Q, симметричную точке P относительно биссектрисы угла xOy. Как мы заметили раньше, точка Q будет иметь координаты $(b;\ a)$. Так как $a = g(b)$, то точка P принадлежит графику функции $y = g(x)$; действительно, при $x = b$ значение $y = a$ равно $g(b)$. Таким образом, все точки, симметричные точкам кривой C относительно указанной прямой, лежат на графике функции $y = g(x)$. Они исчерпывают этот график целиком, так как аналогично показывается, что всякая точка функции $y = g(x)$ при указанной симметрии попадает на график функции $y = f(x)$. Теорема доказана.

Свойства взаимно обратных функций

1. *Тождества*

Пусть f и g — взаимно обратные функции. Это означает, что равенства $y = f(x)$ и $x = g(y)$ равносильны.

Подставим одно из этих равенств в другое. Получим два тождества:

$$f(g(y)) = y \text{ и } g(f(x)) = x.$$

Пример

Функции $y = x^2$, $x \geq 0$ и $y = \sqrt{x}$ взаимно обратны. Имеем два тождества:

$$\left(\sqrt{x}\right)^2 = x \text{ и } \sqrt{x^2} = x \text{ при } x \geq 0.$$

2. *Область определения*

Пусть f и g — взаимно обратные функции. Область определения функции f совпадает с областью значений функции g, и, наоборот, область значений функции f совпадает с областью определения функции g (рис. 36).

Действительно, обратная функция к функции $y = f(x)$ определена для всякого числа y, которое является значением функции для некоторого числа x: мы берем равенство $y = f(x)$ и из него выражаем x как функцию от y. Это свойство наглядно проявляется на графике: график функции $y = f(x)$ совпадает с графиком обратной функции $x = g(y)$, только аргумент функции g откладывается по оси y. Ясно, что аргументы функции g — это значения функции f и наоборот.

Монотонность

Если одна из взаимно обратных функций строго возрастает, то и другая строго возрастает (рис. 37).

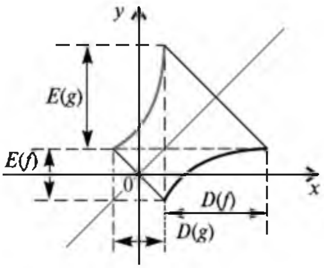


Рис. 36

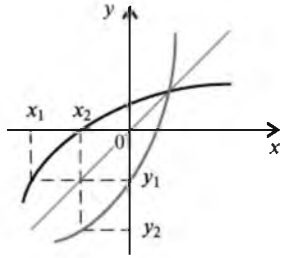


Рис. 37

Действительно, пусть f и g — взаимно обратные функции, причем f строго возрастает. Докажем, что тогда и g строго возрастает. Пусть x_1 и x_2 — два числа, лежащие в области определения функции g, причем $x_1 < x_2$. Надо доказать, что $g(x_1) < g(x_2)$. Обозначим $g(x_1) = y_1$, $g(x_2) = y_2$. Числа y_1 и y_2 лежат в области определения функции f, так как они являются значениями функции g. Предположим, что $y_1 \geq y_2$. В силу монотонности функции f имеем $f(y_1) \geq f(y_2)$. Но $f(y_1) = f(g(x_1)) = x_1$ и $f(y_2) = x_2$, т.е. $x_1 \geq x_2$, что противоречит условию $x_1 < x_2$. Следовательно, $y_1 < y_2$, что и требовалось доказать. Аналогично рассматривается случай убывающей функции (рис. 38).

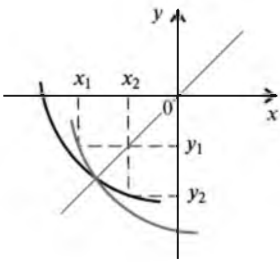


Рис. 38

Условие существования обратной функции

Дана функция $y = f(x)$. Поставим следующий вопрос: при каком условии существует функция, обратная к функции f? По определению, обратная функция к функции f строится так: из соотношения $y = f(x)$ переменную x можно однозначно выразить через y. Мы уже знакомы с примерами функций, для которых это можно сделать. Приведем примеры таких функций, для которых нельзя однозначно выразить аргумент через заданное значение функции.

Примеры

1. $y = |x|$. Для данного положительного числа y найдутся два значения аргумента x такие, что $|x| = y$. Например, если $y = 2$, то $x = 2$ или $x = -2$. Значит, выразить однозначно x через y нельзя.

2. $y = x^2$. Ситуация здесь такая же, как в предыдущем примере: $x = \sqrt{y}$ или $x = -\sqrt{y}$.
3. $y = \sin x$.

Сравним графически эти примеры с примерами, приведенными ранее. Возьмем число y_0 из области значений функции f и проведем прямую $y = y_0$, параллельную оси абсцисс. В ранее рассмотренных случаях эта прямая пересекает график в одной точке, т.е. можно по заданному значению y однозначно найти значение x. В последних трех примерах при некоторых y прямая пересекает график более чем в одной точке: для этих значений y мы не можем однозначно найти x, значит, эти функции не имеют обратных (рис. 39).

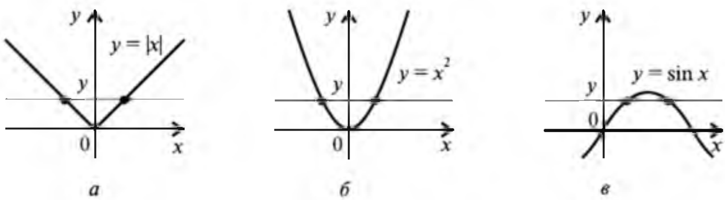


Рис. 39

Дадим геометрическую трактовку условия того, что функция имеет обратную.

Функция $y = f(x)$ имеет обратную, если всякая прямая $y = y_0$ пересекает график функции $y = f(x)$ не более чем в одной точке (она может совсем не пересекать график, если y_0 не принадлежит области значений функции f).

Это же условие можно сформулировать иначе: уравнение $f(x) = y_0$ при каждом y_0 имеет не более одного решения.

Условие того, что функция имеет обратную, заведомо выполняется, если функция строго возрастает или строго убывает. Действительно, если f, например, строго возрастает, то при двух различных значениях аргумента она принимает различные значения, так как большему значению аргумента соответствует большее значение функции. Следовательно, уравнение $f(x) = y$ для строго монотонной функции имеет не более одного решения.

Мы знаем, что многие функции не имеют обратных. Если при некотором b уравнение $f(x) = b$ имеет более одного решения, то функция $y = f(x)$ обратной не имеет. На графике это означает, что прямая $y = b$ пересекает график функции более чем в одной точке.

Многие изучавшиеся ранее функции не имеют обратных, например $y = x^2$. С неоднозначностью решения уравнения $f(x) = b$ можно справиться следующим образом: уменьшить область определения функции f так, чтобы ее область значений не изменилась, но чтобы каждое свое значение она принимала ровно один раз.

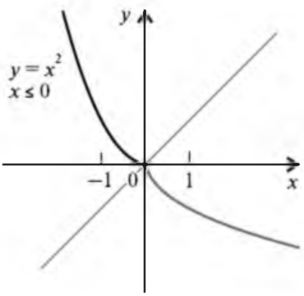


Рис. 40

Примеры

1. $y = x^2, x \ge 0.$
2. $y = x^2, x \le 0.$

В каждом из приведенных примеров функция сохранила область значений: это промежуток $[0; +\infty)$. В то же время, уменьшив область определения, мы добились монотонности функции, а значит, и единственности решения уравнения $f(x) = b$.

Каждая из описанных функций имеет обратную функцию: $y = \sqrt{x}$ для первого примера и $y = -\sqrt{x}$ для второго (рис. 40).

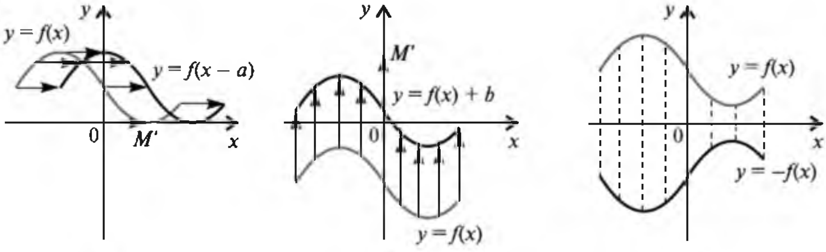
Преобразования графиков

1. ***Параллельный перенос***

Главное назначение понятия функции состоит в том, чтобы с его помощью научиться описывать зависимости между переменными величинами. Основные свойства таких зависимостей не должны существенно меняться при изменении способа измерения этих величин, т.е. при изменении их масштаба и начала отсчета. С другой стороны, за счет более рационального выбора способа измерения переменных величин обычно удается упростить запись зависимости между ними, привести эту запись к некоторому стандартному виду. На геометрическом языке изменение способа измерения величин означает некоторые простые преобразования графиков, к изучению которых мы и переходим.

Изменение начала отсчета переменных приводит к параллельному переносу. Возьмем переменную x и перенесем начало ее отсчета в точку x_0. Новую переменную обозначим x'. Связь между переменными x и x' записывается формулой $x' = x - x_0$ или $x = x' + x_0$. Чтобы не ошибиться в знаке, полезно убедиться в том, что значению исходной переменной $x = x_0$ соответствует нулевое значение переменной x'. Аналогично, сдвигая на y_0 значения переменной y, получаем переменную y', связанную с y формулами $y' = y - y_0$, $y = y' + y_0$.

Установим связь между графиками функций $y = f(x)$ и $y = f(x - a)$. Если мы обозначим $x - a$ через x' (т.е. если мы сдвинем начало отсчета аргумента в точку a, то получим соотношение $y = f(x')$. Это означает, что для построения графика функции $y = f(x - a)$ надо изобразить график исходной функции в системе координат (x', y), т.е. сдвинуть график функции $y = f(x)$ на вектор $\overline{OM'}$ $(a, 0)$. Так как переменная y не меняется, то сдвиг происходит параллельно оси x (рис. 41, a). Этим свойством мы уже пользовались при сравнении графиков функции $y = kx$ и $y = kx + b$.



а *б* *в*

Рис. 41

Аналогично, график функции $y = f(x) + b$ связан с графиком функции $y = f(x)$. Обозначив $y - b$ через y', получим $y' = f(x)$. Это означает, что для построения графика функции $y = f(x) + b$ надо изобразить график исходной функции f в системе координат (x, y'), т.е. сдвинуть график функции $y = f(x)$ на вектор $\overline{OM'}$ $(0, b)$. При этом происходит параллельный перенос вдоль оси y (рис. 41, *б*).

Иногда приходится менять начало отсчета одновременно и у аргумента x, и у функции y, т.е. рассматривать зависимость вида $y = f(x - a) + b$ (или $y - b = f(x - a)$). При построении графика функции $y = f(x - a) + b$ надо сделать два параллельных переноса на векторы $\overline{OA}$ и $\overline{OB}$, которые можно заменить одним параллельным переносом на вектор $\overline{OC}$ (рис. 42). При этом точка (x_0, y_0) графика функции $y = f(x)$ переходит в точку $(x_0 + a, y_0 + b)$ графика функции $y = f(x - a) + b$.

Примеры поведения графика при параллельном переносе приведены на рис. 43, 44.

2. ***Изменение масштаба***

Изменим масштаб измерения величины x. Результат измерения в новом масштабе обозначим через x'. Чтобы найти связь между значениями x и x', достаточно знать, какое значение переменной x' соответствует единице масштаба переменной x. Пусть это значение равно k. Тогда все другие значения изменятся пропорционально, т.е. $\frac{x'}{x} = k$, или $x' = kx$ (проверим, что при $x = 1$ значение x' равно k).

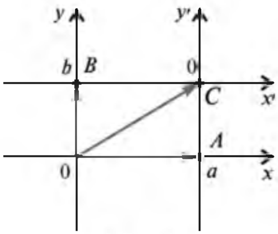


Рис. 42

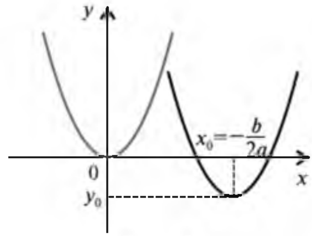


Рис. 43

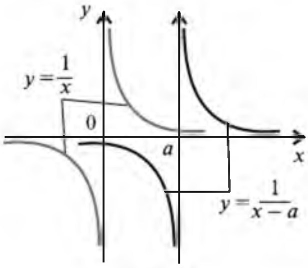


Рис. 44

Отсюда следует связь между графиками функций $y = f(x)$ и $y = f(kx) = f(x')$, где $x' = kx$: график функции $y = f(kx)$ получается из графика функции $y = f(x)$ некоторым растяжением или сжатием вдоль оси x. При $k > 1$ происходит сжатие графика, так как $x' = 1$ получается при $x = \dfrac{1}{k} < 1$ и точка на исходном графике, соответствующая аргументу, равному 1, соответствует точке нового графика при $x = \dfrac{1}{k}$. Аналогично, при $k < 1$ происходит растяжение графика.

Связь между графиками функций $y = f(x)$ и $y = kf(x)$ устанавливается аналогично. Только теперь надо менять масштаб измерения y: $y' = \dfrac{1}{k}\,y$, $y = ky'$. Ситуация стала противоположной: при $k > 1$ происходит растяжение графика вдоль оси y, а при $k < 1$ — его сжатие.

Поведение графика при изменении масштаба изображено на рис. 45.

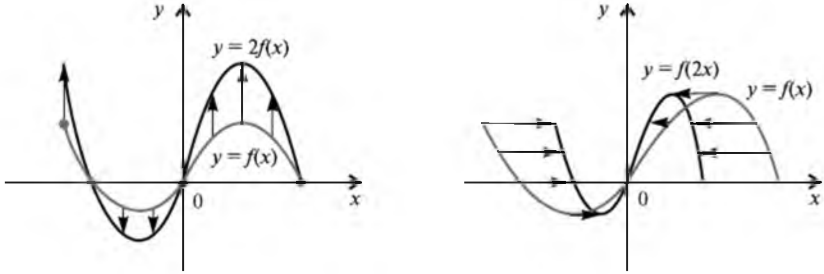


Рис. 45

В преобразованиях $y = f(kx)$ и $y = kf(x)$ мы считали, что $k > 0$. Чтобы включить и случай $k < 0$, рассмотрим сначала $k = -1$. Так как точки (x, y) и $(-x, y)$ симметричны относительно оси y, то и графики функций $y = f(x)$ и $y = f(-x)$ симметричны относительно этой оси (рис. 46). Аналогично, графики функций $y = f(x)$ и $y = -f(x)$ симметричны относительно оси x.

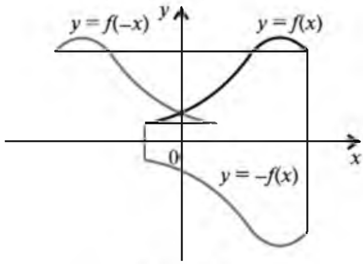


Рис. 46

Пользуясь тремя типами преобразований графиков (параллельный перенос, растяжение — сжатие и симметрия), можно, исходя из графика функции $y = f(x)$, построить график функции $y = Af(kx + b) + B$ при любых значениях параметров A, B, k, b.

Пример

Построить график функции $y = \frac{x}{x+2}$.

Преобразуем правую часть: $\frac{x}{x+2} = \frac{x+2-2}{x+2} = 1 - \frac{2}{x+2}$, т.е. $y - 1 = \frac{-2}{x+2}$, или $y - 1 = \frac{-2}{x+2}$. Отсюда ясно, какие преобразования надо сделать с известным графиком $y = \frac{1}{x}$, чтобы построить требуемый график: его надо отразить относительно оси x (или оси y; это в данном случае безразлично), растянуть по оси y и затем сделать параллельный перенос в точку $(-2, 1)$. Практически поступают так: строят «крест» — прямые $x = -2$ и $y = 1$ — через точку, в которую нужно перенести начало координат, а затем гиперболу располагают в первом — третьем или втором — четвертом квадрантах в зависимости от знака числителя c в записи $y = \frac{c}{x-a} + b$ (в нашем случае $c = -2$) (рис. 47).

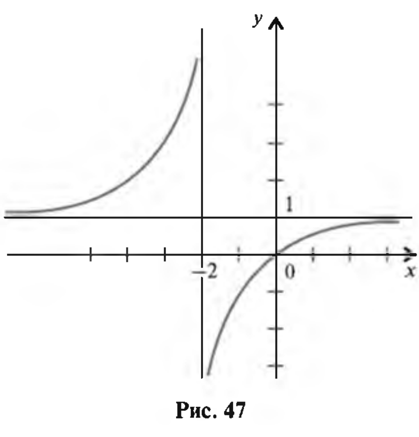


Рис. 47

Контрольные вопросы и задания

1. Проверьте, понимаете ли вы смысл следующих ключевых слов: арифметические операции над функциями, сложная функция, обратная функция, преобразование графиков.
2. Как определяется рациональная функция?
3. Как определяется сложная функция?
4. Расскажите, как связаны между собой свойства и графики взаимно обратных функций.
5. Проверьте, понимаете ли вы смысл обозначений: $f(x-a)$, $f(x)+b$, $f(kx)$.
6. Как перемещается график функции $y=f(x-a)$ при изменении параметра a?
7. Как перемещается график функции $y=f(x)+b$ при изменении параметра b?
8. Как связаны между собой графики функций $y=f(x)$, $y=f(-x)$ и $y=-f(x)$?
9. Как связаны между собой области определения функций $y=f(x)$, $y=f(x-a)$, $y=f(x)+b$, $y=f(-x)$ и $y=-f(x)$?
10. Как изменится график функции $y=f(kx)$ при изменении параметра k? Тот же вопрос для функции $y=kf(x)$.

§ 6. Обзор свойств известных функций

Линейная функция

1. ***Свойства линейной функции***

Линейной функцией называется функция вида $y=kx+b$.

Здесь x — независимая переменная, принимающая произвольные значения (аргумент), y — функция, k и b — параметры. Если $k=0$, то линейная функция становится постоянной ($y=b$). В дальнейшем мы часто будем считать, что $k \neq 0$.

Прямо пропорциональная зависимость между переменными x и y: $\left(\frac{y}{x}=k\right)$ приводит к простейшей линейной функции $y=kx$.

Свойства линейной функции $y=kx$ при $k \neq 0$ (рис. 48).

1. Область определения — множество всех вещественных чисел $\mathbb{R}$.
2. Корни — единственный корень $x=0$.

3. Промежутки постоянного знака. Они зависят от знака параметра k:
   а) $k > 0$: $y > 0$ при $x > 0$; $y < 0$ при $x < 0$;
   б) $k < 0$: $y > 0$ при $x < 0$, $y < 0$ при $x > 0$.
4. Экстремумов нет.
5. Промежутки монотонности:
   а) $k > 0$: y возрастает на всей числовой оси;
   б) $k < 0$: y убывает на всей числовой оси.
6. Наибольших и наименьших значений нет.
7. Область значений — множество $\mathbb{R}$. (Действительно, уравнение $kx = a$ имеет решение при любом a; следовательно, выражение kx ($k \neq 0$) принимает любое значение.)
8. Четность — функция $y = kx$ нечетна.

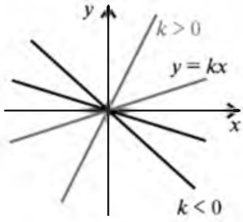


Рис. 48

Свойства функции $y = kx + b$ ($k \neq 0$) могут быть выведены из свойств простейшей линейной функции $y = kx$ с помощью следующего преобразования:

$$kx + b = k\left(x + \frac{b}{k}\right) = k(x - x_1),$$

где $\quad x_1 = -\dfrac{b}{k}$.

Точки x и $x - x_1$ на числовой оси получаются друг из друга сдвигом на $-\dfrac{b}{k}$ (рис. 49), поэтому графики функций $y = kx$ и $y = k(x - x_1)$ получаются друг из друга таким же сдвигом вдоль оси абсцисс. Корень функции «переедет» из точки $x = 0$ в точку $x = x_1 = -\dfrac{b}{k}$. Аналогичный сдвиг произойдет с промежутками постоянного знака и монотонности. Явление нечетности пропадет, хотя симметрия графика сохранится (теперь центром симметрии будет не $x = 0$, а $x = x_1$).

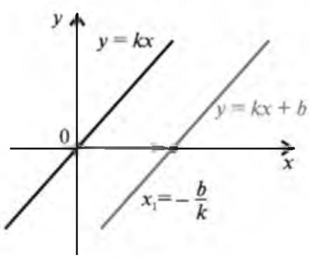


Рис. 49

2. *График линейной функции*

Графиком линейной функции $y = kx$ (прямо пропорциональной зависимости между переменными x и y) является прямая, проходящая через начало координат. Коэффициент k называется *угловым*

коэффициентом этой прямой. Он равен тангенсу угла наклона этой прямой к оси x: $k = \operatorname{tg} \alpha$. При положительных k этот угол острый, при отрицательных — тупой. Это соответствует характеру монотонности линейной функции: при $k > 0$ она возрастает, при $k < 0$ убывает.

Графиком функции $y = kx + b$ является прямая, параллельная прямой $y = kx$, сдвинутая вдоль оси x на $x_1 = -\dfrac{b}{k}$ $(k \neq 0)$ (рис. 50).

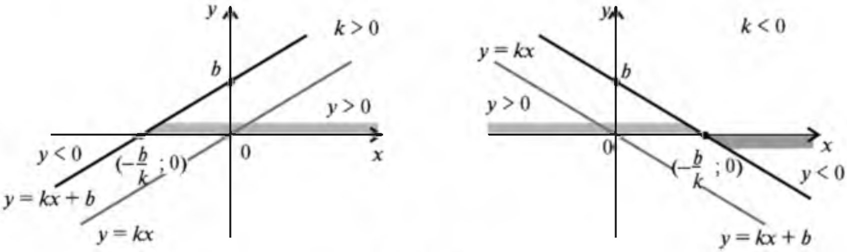


Рис. 50

Для построения графика линейной функции $y = kx + b$ достаточно знать угловой коэффициент k и одну точку, лежащую на графике, т.е. знать значение функции при одном значении аргумента, например $y = y_0$ при $x = x_0$. При необходимости из этих данных легко найти коэффициент b: $y_0 = kx_0 + b$, $b = y_0 - kx_0$.

Итак, прямая, имеющая угловой коэффициент k и проходящая через точку (x_0, y_0), является графиком линейной функции $y = kx + b = kx + (y_0 - kx_0)$. Зависимость $y = kx + (y_0 - kx_0)$ можно переписать в виде $y - y_0 = kx - kx_0$, или $y - y_0 = k(x - x_0)$. Последнюю запись называют уравнением прямой с угловым коэффициентом k, проходящей через точку $(x_0; y_0)$.

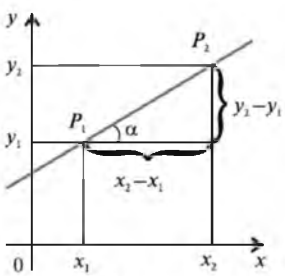


Рис. 51

Часто прямая задается двумя ее точками: $P_1(x_1; y_1)$ и $P_2(x_2; y_2)$. Угловой коэффициент такой прямой вычисляется по формуле $k = \dfrac{y_2 - y_1}{x_2 - x_1}$ (рис. 51). Необходимое для применимости этой формулы условие $x_1 \neq x_2$ геометрически означает, что прямая P_1P_2 не параллельна оси y.

Прямые, заданные различными уравнениями, изображены на рис. 52.

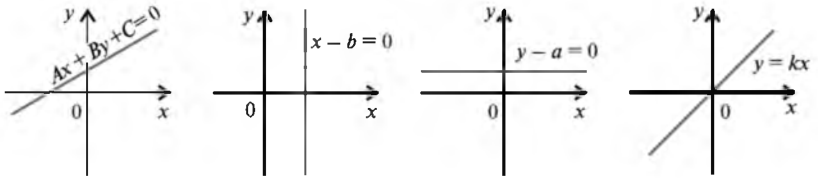


Рис. 52

Квадратичная функция

Мы получим свойства произвольной квадратичной функции, исходя из свойств простейшей функции $y = x^2$.

Пусть $y = ax^2 + bx + c$, $a \neq 0$. Выполним преобразования (выделение полного квадрата из квадратичного трехчлена):

$$ax^2 + bx + c = a\left(x^2 + \frac{b}{a}x + \frac{c}{a}\right) = a\left[x^2 + 2\frac{b}{2a}x + \frac{b^2}{4a^2} - \frac{b^2}{4a^2} + \frac{c}{a}\right] =$$

$$= a\left[\left(x + \frac{b}{2a}\right)^2 - \frac{b^2 - 4ac}{4a^2}\right].$$

Возможные графики при $a > 0$, $D \neq 0$ изображены на рис. 53 (D — дискриминант выражения $ax^2 + bx + c$).

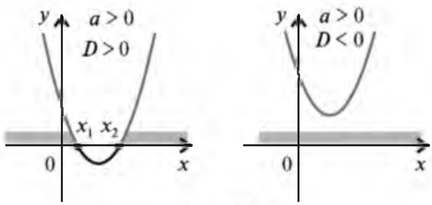


Рис. 53

Сравним графики функций $y = x^2$ и $y = ax^2 + bx + c$ (табл. 1).

Промежутки постоянного знака и промежутки монотонности зависят от знаков чисел a и $D = b^2 - 4ac$. Некоторые случаи изображены на рис. 54.

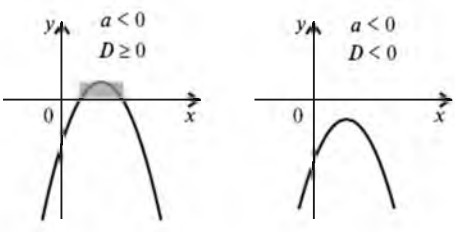


Рис. 54

Таблица 1

Сравнение графиков функций

Функция	$y = x^2$	$y = ax^2 + bx + c$
Вершина	0; 0	$\left(-\frac{b}{2a}; -\frac{b^2-4ac}{4a}\right)$
Корни	$x = 0$	$x = -\frac{b}{2a} \pm \frac{\sqrt{b^2-4ac}}{2a}$ при $b^2 - 4ac \geq 0$, нет корней при $b^2 - 4ac < 0$
Экстремумы	Минимум в вершине	Минимум в вершине, если $a > 0$ Максимум в вершине, если $a < 0$
Область значений	$[0; +\infty)$	$\left[-\frac{b^2-4ac}{4a}; +\infty\right)$, если $a > 0$ $\left(-\infty; -\frac{b^2-4ac}{4a}\right]$, если $a < 0$

Рациональная функция

Рациональной функцией называется функция, значения которой могут быть вычислены по формуле $y = \frac{P(x)}{Q(x)}$, где $P(x)$ и $Q(x)$ — многочлены.

В задании рациональной функции дробью $\frac{P(x)}{Q(x)}$ знаменатель $Q(x)$ может отсутствовать (равняться единице). Это означает, что функции, задаваемые многочленами, мы тоже включаем в число рациональных. Когда хотят выделить случай многочленов, употребляют термин *целые рациональные функции*.

Частный случай рациональной функции — это *дробно-линейная функция*. Так называется функция вида $y = \frac{ax+b}{cx+b}$, которая задается дробью, являющейся отношением двух линейных выражений. Со стандартной дробно-линейной функцией $y = \frac{k}{x}$ мы знакомы. С ее помощью можно задавать обратно пропорциональную зависимость между переменными x и y. Графиком этой функции является гипербола.

Область определения. Рациональная функция, заданная формулой $y = \frac{P(x)}{Q(x)}$, определена во всех точках, кроме тех, при которых знаменатель $Q(x)$ обращается в нуль. Чтобы найти область определения рациональной функции, надо решить уравнение $Q(x) = 0$ и исключить из множества $\mathbb{R}$ его корни.

Обращение в нуль. Рациональная функция $y = \dfrac{P(x)}{Q(x)}$ обращается в нуль там, где $P(x) = 0$. При этом надо исключить «посторонние корни» — те корни числителя, при которых обратился в нуль знаменатель. При таком значении x функция вообще не определена.

Знаки функции. Решение рациональных неравенств вида $\dfrac{P(x)}{Q(x)} > 0$ будет рассмотрено далее.

Исследование монотонности рациональной функции, нахождение наибольших и наименьших значений и построение ее графика требует в общем виде новых, пока не известных нам средств. Полностью это будет доступно при дальнейшем обучении, а сейчас мы ограничимся рассмотрением примеров.

Примеры и комментарии

1. Изучим функцию $y = \dfrac{x}{x+1}$.

Делаем преобразования: $\dfrac{x}{x+1} = \dfrac{x+1-1}{x+1} = \dfrac{x+1}{x+1} - \dfrac{1}{x+1} = 1 - \dfrac{1}{x+1}$.

Запишем результат: $y - 1 = -\dfrac{1}{x-(-1)}$.

Для построения графика строим центр гиперболы — точку $A(-1; 1)$ и проводим через него две прямые, параллельные осям координат. Знак минус перед дробью указывает, что гипербола расположится во 2-й и 4-й четвертях. Осталось найти точки пересечения с осями координат: $x = 0$; $y = 0$ — такая точка одна (рис. 55).

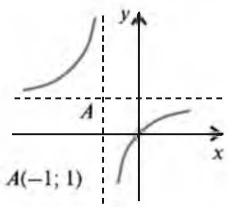


Рис. 55

Результат исследования функции можно записать так:

Область определения: $x \neq -1$.

Нули функции: $x = 0$.

Знаки функции:

$(-\infty; -1)$	$(-1; 0)$	$(0; +\infty)$
+	−	+

Наибольшее и наименьшее значения: нет.

Область значений: $y \neq 1$; ее можно записать так: $(-\infty; 1) \cup (1; +\infty)$.

2. Функция $y = x + \frac{1}{x}$.

Эта функция может быть исследована элементарно, без применения средств математического анализа (производной), которые обычно применяются при построении графиков элементарных рациональных функций.

Область определения: $x \neq 0$.

Нули функции: $x + \frac{1}{x} = 0 \Rightarrow x^2 + 1 = 0$, корней нет.

Знаки функции: при $x > 0$ $y > 0$, при $x < 0$ $y < 0$. Заметим, что если мы поменяем знак у x, то поменяется знак и у y: $y(-x) = -x + \frac{1}{-x} = -\left(x + \frac{1}{x}\right) = -y(x)$. Функция с таким свойством называется *нечетной*. График нечетной функции симметричен относительно начала координат (действительно, точки $P(x; y)$ и $P'(-x; -y)$ одновременно принадлежат или не принадлежат графику). Поэтому достаточно исследовать функцию при $x > 0$. При исследовании любой рациональной функции полезно подумать о ее поведении вблизи ее «особых точек» — так называют корни знаменателя, где функция не определена.

В нашем примере такая точка одна: $x = 0$. Ясно, что при значениях x, близких к нулю, дробь $\frac{1}{x}$ $\left(\text{и, конечно, сумма } x + \frac{1}{x}\right)$ будет принимать большие по модулю значения. Мы будем говорить, что при приближении x к нулю y стремится (приближается) к бесконечности. Важно обратить внимание на знак: если $x > 0$, то $\frac{1}{x} + x > 0$, если же $x < 0$, то $\frac{1}{x} + x < 0$. Можно говорить о приближении x к нулю справа (при $x > 0$) и слева (при $x < 0$). Тем самым можно уточнить: если x приближается к нулю справа, то y стремится к «плюс бесконечности», если же слева, то к «минус бесконечности». На графике такое поведение функции вблизи особой точки $x = 0$ отразится наличием двух «хвостов», приближающихся к прямой $x = 0$ (рис. 56).

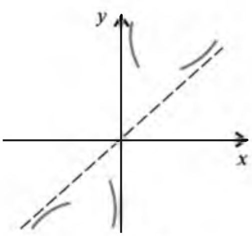


Рис. 56

Теперь посмотрим, как ведет себя функция «на бесконечности», т.е. при больших по модулю значениях x. Ясно, что тогда $\frac{1}{x}$ станет маленьким числом, а сумма $x + \frac{1}{x}$ будет близка к x. На графике это отразится так: график функции $y = x + \dfrac{1}{x}$ будет приближаться к прямой $y = x$. Вообще прямая, к которой приближается график функции при удалении от начала координат, называется *асимптотой*. Теперь мы скажем, что у функции $y = x + \dfrac{1}{x}$ есть две асимптоты — вертикальная $x = 0$ и наклонная $y = x$.

Пытаясь соединить точки при x, близких к нулю, и при больших значениях x, мы получим кривую, изображенную на рис. 57. Осталось уточнить расположение точки, в которой функция принимает наименьшее значение (при $x > 0$). Здесь на помощь придет неравенство для среднего арифметического двух чисел:

$$x + \frac{1}{x} \geq 2\sqrt{x \cdot \frac{1}{x}} = 2.$$

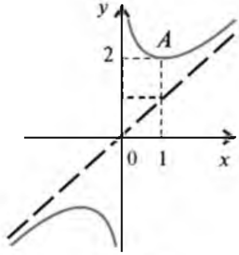


Рис. 57

Можно воспользоваться следствием этого неравенства: если произведение двух положительных чисел x и $\dfrac{1}{x}$ постоянно, то их сумма принимает наименьшее значение, когда слагаемые равны: $x = \dfrac{1}{x} \Rightarrow x = 1$.

Итак, мы нашли нужную нам точку $A(1; 2)$.

График при $x < 0$ строится симметричным отражением относительно начала координат.

Отметим, что построенная кривая также является гиперболой. Ее геометрические свойства мы предложим исследовать в одном из заданий на выбор.

3. Функция $y = \dfrac{x}{x^2 + 1}$.

Область определения: $\mathbb{R}$ (знаменатель в нуль не обращается).

Нули функции: $x = 0$.

Знаки функции: $x > 0 \Rightarrow y > 0$, $x < 0 \Rightarrow y < 0$.

Заметим, что, как и в предыдущем примере, $y(-x) = -y(x)$, т.е. функция нечетная, поэтому достаточно исследовать функцию при $x > 0$.

Изучим *поведение функции на бесконечности*

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

На бесконечности $\frac{1}{x}$ становится маленьким числом, и вся дробь становится близкой к $\frac{1}{x}$. Это означает, что график функции (так же, как график функции $y = \frac{1}{x}$) приближается к оси x (у этой функции есть горизонтальная асимптота $y = 0$).

Так как около нуля дробь $\frac{x}{x^2+1}$ близка к x (ведь в знаменателе к 1 прибавляется ничтожно малая величина x^2), то осталось соединить часть графика около нуля (похожую на кусок прямой $y = x$) с частью графика на бесконечности (похожую на график функции $y = \frac{1}{x}$). Мы опять неизбежно натыкаемся на «горб» — точку, где функция принимает наибольшее значение. Как и в предыдущем примере, нам помогает неравенство для среднего арифметического:

$$x + \frac{1}{x} \geq 2 \Rightarrow \frac{1}{x+\frac{1}{x}} \leq \frac{1}{2} \quad (x > 0).$$

Итак, координаты «горба» графика: $A\left(1; \frac{1}{2}\right)$. Осталось провести плавную кривую (рис. 58).

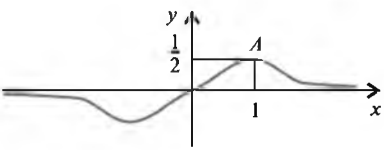


Рис. 58

Функция имеет наименьшее и наибольшее значение ($y = -\frac{1}{2}$ при $x = -1$ и $y = \frac{1}{2}$ при $x = 1$) и ее областью значений является промежуток $\left[-\frac{1}{2}; \frac{1}{2}\right]$.

4. Функция $y = \frac{1}{x^2-1}$.

Область определения: $x \neq \pm 1$ ($x^2 - 1 = 0 \Leftrightarrow x = -1; x = 1$).

Нули функции: нет.

Знаки функции:

$(-\infty; -1)$	$(-1; 1)$	$(1; +\infty)$
+	–	+

Исследуем функцию вблизи «особых точек» $x = -1$ и $x = 1$. Проводим вертикальные прямые и рисуем около них «хвосты» графика (учитывая знаки функции слева и справа от этих точек).

На бесконечности дробь $\frac{1}{x^2-1}$ близка к нулю (а график похож на график функции $y=\frac{1}{x^2}$, с которой мы еще встретимся далее).

Обратим внимание на симметрию функции. Исследуемая функция четна: $y(-x) = y(x)$, поэтому график будет симметричен относительно оси ординат.

Осталось уточнить положение точки максимума между $x = -1$ и $x = 1$. Из соображений симметрии ясно, что этой точкой будет точка с абсциссой $x = 0$. Строим эту точку $P(0; -1)$ и проводим плавную кривую (рис. 59).

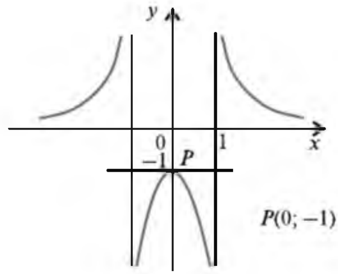


Рис. 59

Областью значений этой функции является множество чисел $[-\infty; -1] \cup \cup (0; +\infty)$.

Контрольные вопросы и задания

1. Проверьте, понимаете ли вы смысл следующих ключевых слов и обозначений: линейная функция, уравнение прямой, угловой коэффициент прямой, $y = kx + b$. Запишите самостоятельно результаты исследования функции $y = kx + b = k(x - x_1)$.
2. Можно ли сказать, что линейная функция является монотонной на всей числовой оси? От чего зависит характер монотонности линейной функции?
3. Сколько раз меняет знак линейная функция? Как определить точку, в которой линейная функция меняет знак?
4. Как вычислить угловой коэффициент прямой, если известны две ее точки?
5. Как вычислить координаты вершины параболы $y = ax^2 + bx + c$?
6. От чего зависит, будут ли ветви параболы $y = ax^2 + bx + c$ смотреть вверх или вниз?
7. Как расположена парабола по отношению к оси Ox, если соответствующая ей квадратичная функция не имеет нулей?
8. Назовите координаты вершины параболы $y = a(x + n)^2$.
9. Назовите формулу дробно-линейной функции. Какая кривая является ее графиком?
10. Что такое асимптота?

§ 7. Использование свойств функций при решении уравнений и неравенств

Решение уравнений и неравенств с помощью графика

Если уравнение записано в виде $f(x) = 0$, то его решения — это корни функции f. Начертив график функции f и отметив его точки пересечения с осью абсцисс, можно приближенно найти решения уравнения, ответить на некоторые качественные вопросы о них, например, сколько корней имеет уравнение на заданном промежутке.

Корни уравнения $f(x) = a$ мы найдем, определив точки пересечения прямой $y = a$, параллельной оси абсцисс, и графика функции f. Это позволяет исследовать количество корней уравнения $f(x) = a$ при различных значениях a. Так, из рисунка 60 видно, что уравнение $f(x) = 1$ имеет пять корней, $f(x) = 3$ — два корня, а уравнение $f(x) = -4$ — ни одного.

Корни уравнения $f(x) = g(x)$ находятся как абсциссы точек пересечения графиков функций f и g. Например, решая уравнение $x^3 + 2x - 3 = 0$, переходим к уравнению $x^3 = -2x + 3$. Строя графики функций $y = x^3$ и $y = -2x + 3$, мы убеждаемся (рис. 61), что графики пересекаются в одной точке, а именно при $x = 1$.

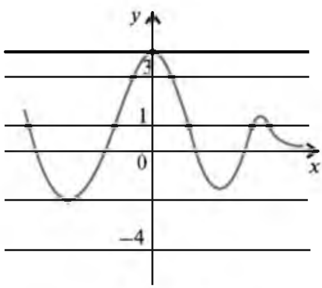


Рис. 60

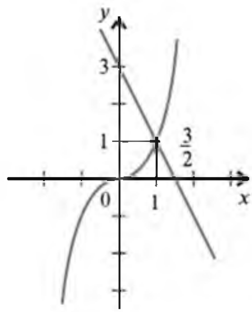


Рис. 61

Решения неравенства $f(x) > 0$ заполняют те промежутки оси, где точки графика функции f лежат выше оси абсцисс. Аналогично, решения неравенства $f(x) > g(x)$ заполняют промежутки, где график функции f лежит выше графика функции g.

Пример

Решить неравенство $x^4 + x^2 - 2 \geq 0$.

Переходим к неравенству $x^4 \geq -x^2 + 2$. Строим графики функций $y = x^4$ и $y = -x^2 + 2$ (рис. 62). Эти графики пересекаются в точках

$P_1(-1;\ 1)$ и $P_2(1;\ 1)$. Решение неравенства: $x \le -1$, $x \ge 1$, т.е. объединение промежутков $(-\infty;\ -1]$ и $[1;\ +\infty)$.

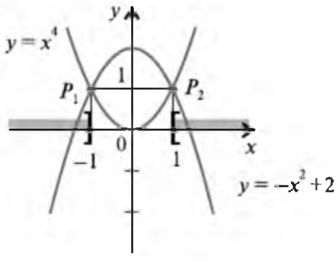


Рис. 62

Графическая иллюстрация решения уравнения указывает, на первый взгляд, и способ решения уравнения: строят в системе координат две кривые и находят их точки пересечения. Действительно, если выбрать масштаб и построить графики достаточно аккуратно, то можно приближенно найти точки пересечения и их абсциссы — корни уравнения. Но для того чтобы найти координаты точек пересечения точно, как раз и нужно решить соответствующее уравнение! В то же время графическая иллюстрация часто дает некоторые качественные ответы, число корней, а также грубо указывает отрезки на числовой оси, где эти корни могут находиться. Рассмотрим в качестве примера уравнение $(x-1)^2 = \sqrt{x}$.

Построим графики функций, стоящих в левой и правой частях.

Из рисунка 63 можно заключить, что уравнение имеет два корня, один из которых находится в интервале $(0;\ 1)$, а другой — в интервале $(2;\ 3)$. Можно указывать эти интервалы и более точно: $(0;\ 0{,}5)$ и $(2;\ 2{,}5)$, еще более точно: $(0{,}2;\ 0{,}3)$ и $(2{,}2;\ 2{,}3)$. (Действительно, нетрудно проверить, что при $x = 0{,}2$ имеем $\sqrt{x} < (x-1)^2$, а при $x = 0{,}3$ – уже $\sqrt{x} > (x-1)^2$; точно так же при $x = 2{,}2$ левая часть уравнения больше правой, а при $x = 2{,}3$ — меньше.) Вообще, вычисляя и сравнивая значения левой и правой частей уравнения, можно найти корни с любой степенью точности.

Корни уравнения пятой степени $x^5 - 3x + 1 = 0$ вообще нельзя записать с помощью радикалов, но, построив достаточно точный график функции $y = x^5 - 3x + 1$ (рис. 64), можно определить, что уравнение имеет три корня в интервалах $(-1{,}5;\ -1{,}3)$, $(0;\ 0{,}5)$ и $(1;\ 1{,}3)$.

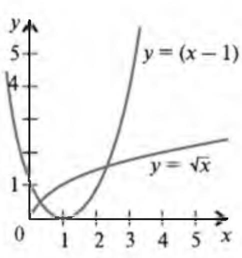


Рис. 63

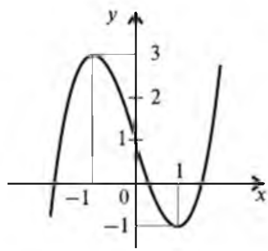


Рис. 64

Метод интервалов

Метод интервалов — это метод решения неравенств вида $f(x) > 0$, когда выражение $f(x)$ можно представить через произведение линейных множителей. Для его применения надо:

1) разложить $f(x)$ на линейные множители;
2) найти корень каждого множителя и нанести все корни на числовую ось;
3) исследовать знак произведения на каждом из получившихся отрезков числовой оси (рис. 65).

Полезно использовать следующее правило. Если все линейные множители различны (имеют разные корни), то произведение изменяет знак при переходе от каждого интервала числовой оси к соседнему (знаки будут чередоваться). Поэтому достаточно определить знак на одном каком-нибудь интервале (обычно это крайний правый интервал). Как применяется это правило, видно из следующих примеров.

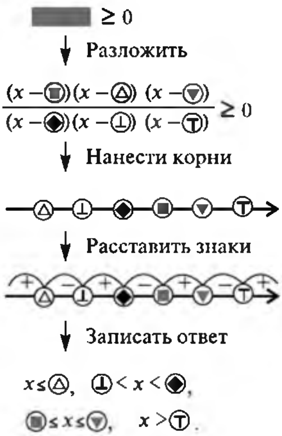


Рис. 65

Примеры

1. Решить неравенство $(x-1)(x-3) > 0$.

Для решения нанесем на числовую ось точки $x_1 = 1$; $x_2 = 3$ (корни линейных функций $y = x - 1$ и $y = x - 3$). Эти точки разбивают числовую ось на три интервала: $(-\infty; 1)$, $(1; 3)$, $(3; +\infty)$. На каждом из интервалов каждый множитель сохраняет постоянный знак, а при переходе через корень меняет знак один множитель. Начнем с крайнего правого интервала $(3; +\infty)$. На нем оба множителя положительны. При переходе (справа налево) через точку $x_2 = 3$ множитель $x - 3$ стал отрицательным и все произведение поменяло знак, стало отрицательным. При переходе через точку $x_1 = 1$ поменял знак первый множитель, и все произведение стало положительным. Результат исследования знаков записывают схемой, показанной на рис. 66. *Ответ:* $(-\infty; 1) \cup (3; +\infty)$, т.е. $x < 1$; $x > 3$.

2. Решить неравенство $\dfrac{(x+1)(x-5)}{x(x-1)} < 0$.

Рис. 66

Нанесем на числовую ось корни линейных функций, с помощью которых образована дробь $\dfrac{(x+1)(x-5)}{x(x-1)}$. Точки -1; 0; 1; 5

разбили числовую ось на пять интервалов. Распределение знаков дроби на этих интервалах изображено на рис. 67. *Ответ:* $(-1; 0) \cup (1; 5)$, т.е. $-1 < x < 0$; $1 < x < 5$.

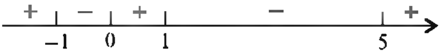


Рис. 67

Иногда для применения метода интервалов приходится раскладывать на множители, выносить числовые множители для того, чтобы записать левую часть в стандартном виде — в виде произведения или частного выражений вида $x - a$.

3. Решить неравенство $\dfrac{x^2+x-2}{5-2x} > 0$.

Сделаем преобразование $\dfrac{x^2+x-2}{5-2x} = \dfrac{(x+2)(x-1)}{2(x-5/2)}$. Получим $\dfrac{(x+2)(x-1)}{x-5/2} < 0$.

Корни линейных функций: $x_1 = -2$; $x_2 = 1$; $x_3 = \dfrac{5}{2}$. *Ответ:* $(-\infty, -2) \cup \cup \left(1, \dfrac{5}{2}\right)$ (рис. 68). При решении нестрогих неравенств вида $f(x) \geq 0$ или $f(x) \leq 0$ надо включать в множество решений точки, являющиеся корнями линейных функций, стоящих в числителе.

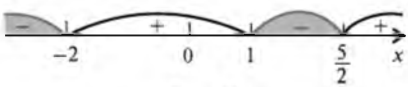


Рис. 68

4. Решить неравенство $\dfrac{x(2-x)}{(2x+5)(2x+3)} \geq 0$.

Перепишем неравенство в равносильной форме:

$$\frac{x(2-x)}{2 \cdot 2(x+5/2)(x+3/2)} \geq 0.$$

Нанесем на числовую ось точки $-\dfrac{5}{2}$; $-\dfrac{3}{2}$; 0; 2. *Ответ:* $\left(-\dfrac{5}{2}; -\dfrac{3}{2}\right) \cup \cup [0; 2]$ (рис. 69).

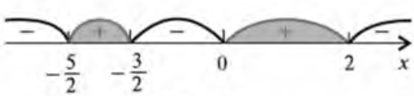


Рис. 69

Существенной чертой метода интервалов является разбиение числовой оси на участки и рассмотрение данной функции отдельно на каждом участке. Это же обычно приходится делать, когда нужно «раскрыть модули».

5. Построить график функции $y = |x^2 - 4| + x^2$.

Выясним, при каких величинах x выполняется неравенство $x^2 - 4 \geq 0$. Решая это неравенство с помощью метода интервалов, получим два луча $x \geq 2$ и $x \leq -2$ (рис. 70). При этих x выражение под знаком модуля неотрицательно, поэтому, раскрывая модуль, получаем, что при этих x $y = x^2 - 4 + x^2 = 2x^2 - 4$.

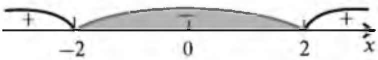


Рис. 70

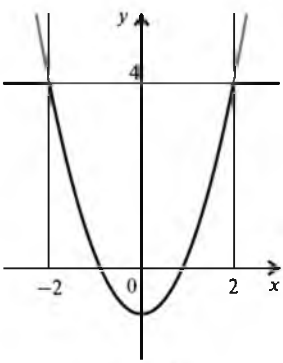


Рис. 71

При остальных x, т.е. при $-2 < x < 2$, выражение под знаком модуля отрицательно. Поэтому $y = -(x^2 - 4) + x^2 = 4$. Итак, на отрезке $-2 < x < 2$ данная функция совпадает с функцией $y = 4$, а на двух лучах $x < -2$ и $x > 2$ — с функцией $y = 2x^2 - 4$ (рис. 71).

При разложении могут встретиться одинаковые множители. Если этих множителей четное число, то при переходе аргумента через их общий корень произведение не изменяет знака. Если число одинаковых множителей нечетно, то знак изменяется. Множители в четной степени, хотя они и не влияют на знак всего выражения, отбрасывать нельзя, так как при этом потеряется точка, в которой этот множитель обращается в нуль.

6. Решить неравенство $\dfrac{(x+1)^3(x-2)^2}{(x-1)^2} \geq 0$.

Рис. 72

Нанесем на числовую ось корни линейных функций $x_1 = -1$, $x_2 = 1$, $x_3 = 2$ и определим знаки всего выражения на каждом промежутке (рис. 72). *Ответ:* $[-1; 1) \cup (1; +\infty)$. Интересно сравнить ответы для неравенств, в которых левая часть одна и та же, а правая меняется:

при > 0 $(-1, 1) \cup (1, 2) \cup (2, +\infty)$;
при ≤ 0 $(-\infty, -1)$, $x = 2$;
при < 0 $(-\infty; -1)$.

Контрольные вопросы и задания

1. Проверьте, правильно ли вы понимаете смысл слов: графическое решение уравнений, графическое решение неравенств, метод интервалов.
2. Как графически решается уравнение: $f(x) = 0$; $f(x) = a$; $f(x) = g(x)$?
3. Как графически решается неравенство: $f(x) > 0$, $f(x) > g(x)$?
4. Как применяется метод интервалов для решения неравенств?
5. Сформулируйте правило для определения знаков произведения различных множителей при решении неравенств методом интервалов.
6. Что происходит со знаками произведения, если не все множители различны?
7. Сформулируйте правило раскрытия модуля выражения.
8. Как применяется метод интервалов для раскрытия модуля выражения?
9. Что такое нестрогое неравенство?
10. Расскажите об особенностях решения нестрогого неравенства.

ГЛАВА 2

КООРДИНАТЫ И ВЕКТОРЫ

§ 8. Направленные отрезки

Примеры векторных величин

Появление многих математических понятий связано с задачей измерения величин. Для таких величин, как длина, температура, работа, масса, в результате измерения получается число, причем в этих случаях это число служит полным описанием измеряемой величины. О таких величинах говорят, что они являются *скалярными*. Значения скалярной величины можно расположить на координатной прямой, шкале.

Понятие «вектор» охватывает более широкий класс величин. В этот класс входят такие величины, как, скажем, сила, скорость, напряженность электрического поля и др.

Вектор — это величина, которая задается своим модулем (длиной) и направлением.

Вектор изображается направленным отрезком, длина которого равна модулю вектора.

Примечание. Вектор обозначается обычно прямой полужирной буквой или буквой, над которой ставится стрелка: **a**, $\vec{F}$, $\overline{AB}$. Модуль вектора обозначается той же светлой буквой или буквой со знаком модуля.

Примеры

1. На рисунке 73 изображены силы, приложенные к саням, которые тянут за веревку.

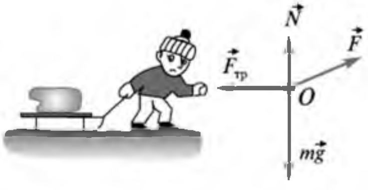


Рис. 73

Точкой O обозначены сани (точка приложения сил), $m\mathbf{g}$ — сила тяжести, **N** — сила нормального давления (реакция опоры), **F** — сила натяжения веревки, $\mathbf{F}_{тр}$ — сила трения.

2. Движение мяча (рис. 74) описывается следующими векторными величинами: **r** — перемещение мяча, **v** — скорость, $m\mathbf{g}$ — сила тяжести, действующая на мяч, ω — угловая скорость вращения мяча вокруг оси.

Рис. 74

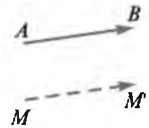


Рис. 75

3. Рассмотрим параллельный перенос пространства. Выберем в пространстве направленный отрезок AB и сопоставим ему параллельный перенос, т.е. преобразование пространства, при котором точка M переходит в такую точку M', что отрезки $\overline{MM'}$ и $\overrightarrow{AB}$ имеют одинаковую длину, параллельны и одинаково направлены. *Параллельный перенос* — это вектор. Он определяется не только длиной отрезка AB, но и его направлением (рис. 75).

Изображение векторов направленными отрезками

Геометрически векторные величины изображают направленными отрезками. Не следует говорить, что вектор — это направленный отрезок. Скажем, ни сила, ни скорость, ни параллельный перенос не являются сами по себе отрезками. Другое дело, что их можно изображать, задавать направленными отрезками, расположенными на плоскости или в пространстве. Сформулируем соответствующие правила.

Правила изображения векторов

1. Однородность

От любой точки можно отложить направленный отрезок, изображающий данный вектор.

2. Равенство векторов

Направленные отрезки изображают один и тот же вектор в том и только в том случае, когда отрезки равны по длине, параллельны и одинаково направлены (рис. 76).

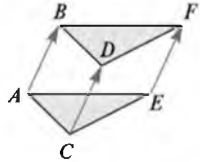


Рис. 76

Эти правила отражают определение вектора как величины, задающейся модулем и направлением. Модуль вектора (его длина, абсолютная величина), равен длине изображающего этот вектор отрезка, измеренной с помощью единицы длины. Сложнее обстоит дело с заданием направления, так как трудно коротко ответить на вопрос, что же такое направление. Направление на прямой задается легко. Мы ставим стрелку в одну из двух возможных сторон (вправо—влево, вверх—вниз и др.). Задав направление на одной прямой, легко задать такое же направление на другой прямой, параллельной исходной. Иными словами, мы умеем наглядно сравнивать выбранные направления на двух параллельных прямых — совпадают эти направления или противоположны.

Поэтому, откладывая один и тот же вектор от разных точек, надо следить за тем, чтобы строящиеся отрезки лежали на параллельных прямых и шли в одинаковом направлении.

Кроме того, что векторы задаются модулем и направлением, они имеют другую важную характеристику. Их можно складывать и умножать на числа.

3. Правило трех точек

Если отрезок AB изображает вектор $\mathbf{a}$, отрезок BC — вектор $\mathbf{b}$, то отрезок AC изображает сумму $\mathbf{a} + \mathbf{b}$ (рис. 77).

4. Растяжение

Если отрезок AB изображает вектор $\mathbf{a}$, то вектор $\lambda\mathbf{a}$ можно изобразить как отрезок AC, лежащий на прямой AB, с длиной $|AC| = |\lambda|\,|AB|$ и с направлением, совпадающим с направлением отрезка AB, если $\lambda > 0$, и противоположным ему, если $\lambda < 0$ (рис. 78).

5. Правило параллелограмма

Пусть $\overline{OA_1} = \mathbf{a}$ и $\overline{AB} = \mathbf{b}$. Тогда диагональ OA_3, параллелограмма $OA_1A_2A_3$ со сторонами OA_1, OA_2 изображает сумму векторов $\mathbf{a}$ и $\mathbf{b}$ (рис. 79).

6. Изображение противоположного вектора

Пусть $\overline{AB} = \mathbf{a}$ и точка C симметрична B относительно A. Тогда отрезки AC и BA изображают вектор $-\mathbf{a}$, направленный противоположно вектору $\mathbf{a}$ (рис. 80).

7. Изображение нулевого вектора

Нулевой вектор $\mathbf{0}$ изображается точкой, т.е. отрезком, у которого начало и конец совпадают (рис. 81).

8. Правило многоугольника

Если несколько векторов изображены так, что начало второго есть конец первого, начало третьего — конец второго и т.д., то отрезок, соединяющий начало первого с концом последнего, изображает сумму этих векторов (рис. 82).

9. Изображение разности

Если два вектора a и b отложены от одной точки O: $\overline{OA}$ = a, $\overline{OB}$ = b, то разность их b – a изображается отрезком $\overline{AB}$, соединяющим концы (рис. 83). **Полезно также запомнить, что диагонали параллелограмма изображают векторную сумму и разность сторон этого параллелограмма.**

10. Правило параллелепипеда

Если $\overline{OA_1}$ = a, $\overline{OA_2}$ = b и $\overline{OA_3}$ = c, то диагональ $\overline{OA}$ параллелепипеда со сторонами OA_1, OA_2, OA_3 изображает сумму векторов a + b + c (рис. 84).

Первые четыре из сформулированных правил являются основными, принимаемыми без доказательства. Остальные правила можно из них вывести.

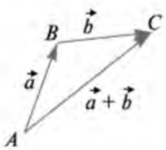
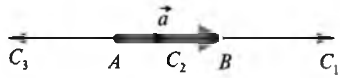


Рис. 77



$$\overrightarrow{AC_1} = 2\overrightarrow{AB}$$

$$\overrightarrow{AC_2} = \frac{1}{2}\overrightarrow{AB}$$

$$\overrightarrow{AC_3} = -\overrightarrow{AB}$$

Рис. 78

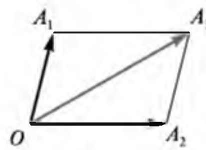
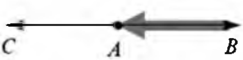
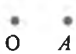


Рис. 79



$\overrightarrow{AC} = -\overrightarrow{AB}$ $\quad$ $\overrightarrow{BA} = -\overrightarrow{AB}$

Рис. 80



$\overrightarrow{AA} = 0$

Рис. 81

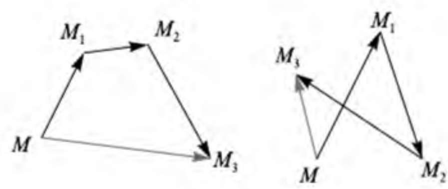


Рис. 82

$\overrightarrow{AB} = \overrightarrow{OB} - \overrightarrow{OA}$

Рис. 83

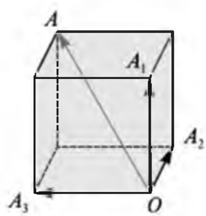


Рис. 84

Докажем, например, *правило параллелограмма*.

Действительно, из свойств параллелограмма и из правил трех точек следует: $\overline{A_1A_3} = \overline{OA_2}$, $\overline{OA_1} + \overline{A_1A_3} = \overline{OA_3}$. Таким образом, $\overline{OA_1} + \overline{OA_2} = \overline{OA_3}$.

Изображение векторов в виде направленных отрезков помогает переносить на векторы различные геометрические понятия, например такие, как угол, расстояние, перпендикулярность и др.

Рис. 85

Как определить, например, угол между двумя векторами? Возьмем два ненулевых вектора. Отложим их от одной и той же точки. Полученный геометрический угол и считается углом между векторами (рис. 85). Разумеется, его величина не зависит от того, от какой точки откладывались векторы. Заметим, что угол между векторами не зависит от их длин, а зависит только от направлений.

Применение векторов к решению геометрических задач

Пусть O — фиксированная точка. *Радиус-вектором точки P* (относительно точки O) называют вектор $\overline{OP}$.

Использование векторов при решении геометрических задач основано на том, что каждую точку можно задавать радиус-вектором (рис. 86).

Выведем формулу для радиуса-вектора середины отрезка.

Теорема. Пусть точка C — середина отрезка AB. Тогда радиус-вектор точки C есть полусумма радиусов-векторов точек A и B, т.е. $\overline{OC} = \frac{1}{2}(\overline{OA} + \overline{OB})$, где O — произвольно выбранная точка (рис. 87).

Доказательство. Согласно правилу трех точек, $\overline{OC} = \overline{OA} + \overline{AC}$, $\overline{OC} = \overline{OB} + \overline{BC}$. Складывая эти равенства, мы получаем $2\overline{OC} = \overline{OA} + \overline{OB} + \overline{AC} + \overline{BC}$, $\overline{OC} = \frac{\overline{OA} + \overline{OB}}{2} + \frac{\overline{AC} + \overline{BC}}{2}$.

Но, поскольку $\overline{AC} = -\overline{BC}$ (правило 4), сумма $\overline{AC} + \overline{BC}$ равна $\mathbf{0}$. Следовательно, $\overline{OC} = \frac{\overline{OA} + \overline{OB}}{2}$.

Примеры

1. Доказать, что середины сторон четырехугольника образуют параллелограмм (рис. 88).

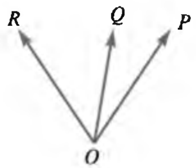


Рис. 86

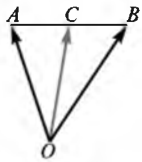


Рис. 87

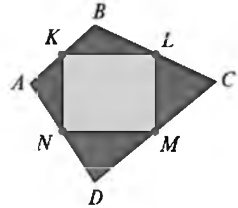


Рис. 88

Пусть K, L, M, N — середины сторон AB, BC, CD, AD четырехугольника $ABCD$. Если $\overline{KL} = \overline{NM}$, то точки K, L, M, N образуют параллелограмм. Докажем, что $\overline{KL} = \overline{NM}$. Согласно теореме, примененной к серединам сторон $\overline{AB}$, $\overline{BC}$, $\overline{CD}$, $\overline{DA}$:

$$\overline{OK} = \frac{1}{2}(\overline{OA} + \overline{OB}),\ \overline{OM} = \frac{1}{2}(\overline{OC} + \overline{OD}),$$

$$\overline{OL} = \frac{1}{2}(\overline{OB} + \overline{OC}),\ \overline{ON} = \frac{1}{2}(\overline{OD} + \overline{OA}).$$

Вычитая из третьего равенства первое и из второго четвертое, мы получаем (правило 9):

$$\overline{KL} = \frac{1}{2}(\overline{OC} - \overline{OA}),\ \overline{NM} = \frac{1}{2}(\overline{OC} - \overline{OA})$$

и

$$\overline{KL} = \overline{NM}.$$

Заметим, что в этой задаче никаких условий на четырехугольник не накладывалось. Он мог быть, например, не плоским или самопересекающимся.

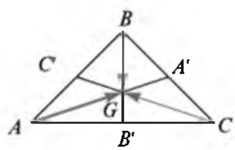


Рис. 89

2. Пусть G — точка пересечения медиан треугольника ABC. Доказать, что $\overline{AG} + \overline{BG} + \overline{CG} = \mathbf{0}$ (рис. 89).

Обозначим A', B', C' середины сторон BC, CA и AB. Согласно известному свойству точки пересечения медиан: $\overline{AG} = \frac{2}{3}\overline{AA'}$, $\overline{BG} = \frac{2}{3}\overline{BB'}$, $\overline{CG} = \frac{2}{3}\overline{CC'}$.

Складывая эти равенства, получаем

$$\overline{AG} + \overline{BG} + \overline{CG} = \frac{2}{3}(AA' + BB' + CC').$$

По теореме о радиус-векторе середины отрезка, имеем:

$$\overline{AA'}=\frac{1}{2}(\overline{AB}+\overline{AC}),\ \overline{BB'}=\frac{1}{2}(\overline{BA}+\overline{BC}),\ \overline{CC'}=\frac{1}{2}(\overline{CA}+\overline{CB}).$$

Таким образом,

$$\overline{AA'}+\overline{BB'}+\overline{CC'}\ =\ \frac{1}{2}(\overline{AB}+\overline{AC}+\overline{BA}+\overline{BC}+\overline{CA}+\overline{CB})=\mathbf{0}\ ,$$

и так как

$$\overline{AG}+\overline{BG}+\overline{CG}=\frac{2}{3}(\overline{AA'}+\overline{BB'}+\overline{CC'})\ ,$$

то

$$\overline{AG}+\overline{BG}+\overline{CG}=\mathbf{0}\ .$$

Полученное в этой задаче равенство можно переписать так: $\overline{GA}+\overline{GB}+\overline{GC}=\mathbf{0}$. Это свойство точки пересечения медиан имеет физический смысл. Оно означает, что в том случае, когда точки A, B, C имеют одинаковые массы, центр масс системы A, B, C находится в точке пересечения медиан треугольника ABC. Дело в том, что центр масс системы точек A, B, C, имеющих массы m_A, m_B, m_C, определяется равенством $m_A\overline{GA}+m_B\overline{GB}+m_C\overline{GC}=\mathbf{0}$.

Проекция вектора

Как найти проекцию вектора **a** на ось l? Изобразим вектор **a** направленным отрезком $\overline{AB}$ и спроектируем его концы на ось l.

Проекцией вектора на ось называется число, равное разности координат проекций на эту ось конца и начала этого вектора.

Иными словами, если **a** = $\overline{AB}$, то проекция вектора **a** на ось l равна числу $B_l - A_l$, где B_l и A_l — координаты проекций точек B и A на ось l. Обозначим проекцию вектора **a** на ось l через a_l. Получим равенство (рис. 90)

$$a_l = B_l - A_l.$$

Если этот же вектор **a** изобразить другим направленным отрезком $\overline{CD}$, то проекция вектора **a** на ось l не изменится, т.е. если $\overline{AB}\ =\ \overline{CD}$, то $D_l - C_l = B_l - A_l$. Это наглядно видно из рис. 91.

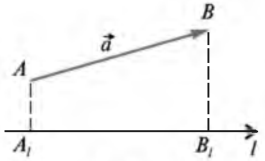


Рис. 90

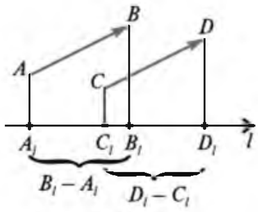


Рис. 91

Свойства проекции вектора на ось

Теорема 1. Проекция вектора на ось равна длине этого вектора, умноженной на косинус угла между вектором и осью.

Пусть φ — угол между вектором a и осью l (рис. 92). Тогда свойство можно записать в виде формулы

$$a_l = |\mathbf{a}| \cos \varphi.$$

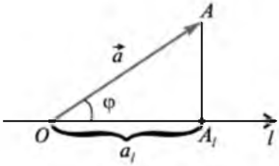


Рис. 92

Доказательство. Отложим вектор **a** от начала O оси l: $\mathbf{a} = \overline{OA}$. Пусть A_l — проекция точки A на ось l. По определению, проекция вектора **a** на ось l равна координате точки A_l, так как координата начала O равна нулю. Координата точки A_l равна длине отрезка OA, умноженной на $\cos \varphi$, т.е. $a_l = |OA| \cos \varphi = |\mathbf{a}| \cos \varphi$, что и требовалось доказать.

Теорема 2. Проекция суммы векторов равна сумме проекций:

$$(\mathbf{a} + \mathbf{b})_l = a_l + b_l.$$

Доказательство. Изобразим вектор **a** направленным отрезком $\overline{AB}$. Отложим вектор **b** от точки B: $\mathbf{b} = \overline{BC}$. Вектор $\overline{AC}$ равен сумме векторов **a** и **b**: $\mathbf{a} + \mathbf{b} = \overline{AC}$. Рассмотрим проекции точек A, B и C на ось l (рис. 93). Получим

$$a_l = B_l - A_l; \quad b_l = C_l - B_l.$$

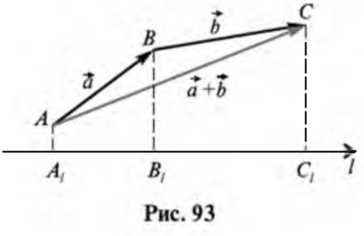


Рис. 93

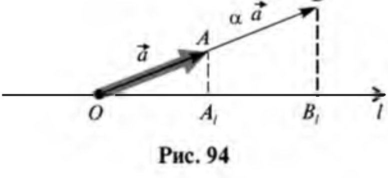


Рис. 94

Сложим эти равенства: $a_l + b_l = B_l - A_l + C_l - B_l = C_l - A_l$. Мы получим проекцию вектора $\overline{AC} = \mathbf{a} + \mathbf{b}$, что и требовалось доказать.

Теорема 3. При умножении вектора на число его проекция тоже умножается на это число:

$$(\alpha\mathbf{a})_l = \alpha a_l.$$

Доказательство. Изобразим вектор **a** направленным отрезком $\overline{OA}$. Отложим вектор $\alpha\mathbf{a}$ от точки O: $\alpha\mathbf{a} = \overline{OB}$ (рис. 94). Найдем проекции точек A и B на ось l. Из свойств пропорциональных отрезков имеем $OB_l = \alpha OA_l$, т.е. проекция вектора $\alpha\mathbf{a} = \overline{OB}$ равна проекции вектора $\mathbf{a} = \overline{OA}$, умноженной на число α, что и требовалось доказать.

Контрольные вопросы и задания

1. Обратите внимание на следующие ключевые выражения, появившиеся в параграфе; векторная величина, угол между векторами, проекция вектора на ось. Приведите примеры их употребления.
2. Приведите примеры векторных и скалярных величин.
3. Когда два направленных отрезка изображают один и тот же вектор?
4. В чем состоит правило трех точек?
5. Сформулируйте правила параллелограмма и параллелепипеда. Докажите самостоятельно правило изображения разности.
6. Как изображается противоположный вектор?
7. Как геометрически складываются несколько векторов?
8. Какой векторный смысл имеют диагонали параллелограмма?
9. Как определяется проекция вектора на ось?
10. Какие вы знаете свойства проекции вектора на ось?

§ 9. Скалярное произведение

Определение скалярного произведения

Скалярным произведением двух векторов называется произведение их длин на косинус угла между ними.

Если обозначить векторы **a** и **b**, а угол между ними — φ, то определение скалярного произведения можно записать в виде формулы

$$\mathbf{a} \cdot \mathbf{b} = |\mathbf{a}| \cdot |\mathbf{b}| \cos \varphi.$$

Знак скалярного умножения обозначается точкой, как и обычное умножение чисел.

1. Если хотя бы один из векторов **a** или **b** нулевой, то угол между векторами не определен. В этом случае, по определению, считается, что скалярное произведение $\mathbf{a} \cdot \mathbf{b}$ равно нулю.
2. Важно помнить, что в результате скалярного умножения двух векторов получается число, скаляр, поэтому произведение и названо скалярным.

Пример

Предположим, что мы тянем за веревку сани с постоянной силой **F** (рис. 95). Какую работу мы совершим, переместив сани на расстояние **s**?

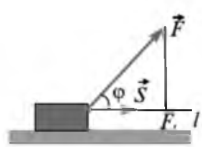


Рис. 95

Если мы разложим силу **F** по горизонтальному и вертикальному направлениям, то станет ясно, что работу по перемещению выполняет лишь горизонтальная составляющая $\mathbf{F}_l$:

$$A = |\mathbf{F}_l| \cdot |\mathbf{s}| = |\mathbf{F}| \cos \varphi \, |\mathbf{s}| = \mathbf{F} \cdot \mathbf{s},$$

где φ — угол между направлениями силы тяги и перемещения.

Формула для скалярного произведения через проекции

Разбирая пример вычисления работы, мы выяснили, что эту работу совершает лишь горизонтальная составляющая силы, т.е. ее проекция на направление перемещения. Поэтому скалярное произведение можно вычислить через проекцию одного вектора на направление другого.

Пусть **a** и **b** — два вектора, причем $\mathbf{b} \neq 0$. Проведем прямую l, параллельную вектору **b**. Положительное направление оси l определяем по направлению **b**. Проекцией вектора **a** на направление вектора **b** называют проекцию **a** на построенную ось l. Обозначают эту проекцию так: a_b. Пусть угол φ — угол между векторами **a** и **b**. Ясно, что это то же самое, что угол между вектором **a** и осью l (рис. 96). Из свойств проекции следует, что $a_b = |\mathbf{a}| \cos \varphi$. Сравнивая эту формулу с определением скалярного произведения, получаем:

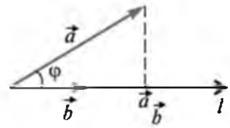


Рис. 96

$$\mathbf{a} \cdot \mathbf{b} = a_b \cdot |\mathbf{b}|.$$

Скалярное произведение двух векторов равно проекции одного из них на направление другого, умноженной на длину этого второго вектора.

В примере механической работы получаем

$$A = \mathbf{F} \cdot \mathbf{s} = F_s \cdot |\mathbf{s}|,$$

где F_s — горизонтальная составляющая силы, т.е. ее проекция на направление вектора $\mathbf{s}$.

Заметим, что мы на самом деле получили две формулы, так как можно проектировать $\mathbf{a}$ на $\mathbf{b}$, а можно, наоборот, $\mathbf{b}$ на $\mathbf{a}$:

$$\mathbf{a} \cdot \mathbf{b} = a_b \cdot |\mathbf{b}| = b_a \cdot |\mathbf{a}|.$$

Отметим еще, что во всех этих формулах надо считать, что векторы $\mathbf{a}$ и $\mathbf{b}$ ненулевые.

Свойства скалярного произведения

Перечислим некоторые свойства скалярного произведения векторов, вытекающие непосредственно из его определения.

Свойство 1

Скалярный квадрат вектора равен квадрату его длины:

$$\mathbf{aa} = \mathbf{a}^2 = |\mathbf{a}|^2.$$

В самом деле,

$$\mathbf{a}^2 = |\mathbf{a}|^2 \cos \angle (\mathbf{a}, \mathbf{a}) = |\mathbf{a}|^2,$$

так как $\angle (\mathbf{a}, \mathbf{a}) = 0$, а $\cos 0 = 1$.

Свойство 2

Скалярное умножение векторов перестановочно:

$$\mathbf{ab} = \mathbf{ba}.$$

В самом деле,

$$\angle (\mathbf{a}, \mathbf{b}) = \angle (\mathbf{b}, \mathbf{a}) \text{ и } \mathbf{ab} = |\mathbf{a}| \cdot |\mathbf{b}| \cos \angle (\mathbf{a}, \mathbf{b}) = \mathbf{ba}.$$

Свойство 3

Скалярное умножение векторов линейно:

$$\mathbf{a}(\mathbf{b} + \mathbf{c}) = \mathbf{a} \cdot \mathbf{b} + \mathbf{a} \cdot \mathbf{c} \text{ и } \mathbf{a}(\alpha\mathbf{b}) = \alpha(\mathbf{a} \cdot \mathbf{b}).$$

Для доказательства воспользуемся сначала формулой для скалярного произведения через проекции, а затем свойствами проекции:

$$\mathbf{a}(\mathbf{b} + \mathbf{c}) = |\mathbf{a}| \cdot (\mathbf{b} + \mathbf{c})_a = |\mathbf{a}| \cdot b_a + |\mathbf{a}| \cdot c_a = \mathbf{a} \cdot \mathbf{b} + \mathbf{a} \cdot \mathbf{c};$$
$$\mathbf{a}(\alpha\mathbf{b}) = |\mathbf{a}| \cdot (\alpha\mathbf{b})_a = |\mathbf{a}|\, \alpha b_a = \alpha|\mathbf{a}| \cdot b_a = \alpha\, (\mathbf{a} \cdot \mathbf{b}).$$

Первые два свойства позволяют обращаться со скалярным умножением векторов, как с умножением чисел. Например:

$$(\mathbf{a}+\mathbf{b})(\mathbf{a}-\mathbf{b}) = \mathbf{a}^2 - \mathbf{b}^2,$$
$$(\mathbf{a}+\mathbf{b})^2 = \mathbf{a}^2 + \mathbf{b}^2 + 2\mathbf{ab},$$
$$(\mathbf{a}-\mathbf{b})^2 = \mathbf{a}^2 + \mathbf{b}^2 - 2\mathbf{ab}.$$

Последнее тождество имеет важные применения в геометрии.

Рассмотрим треугольник ABC (рис. 97). В векторных обозначениях $\overline{BC} = \overline{AC} - \overline{AB}$ и поэтому

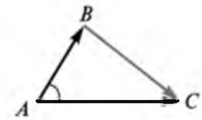


Рис. 97

$$|\overline{BC}|^2 = |\overline{AB}|^2 + |\overline{AC}|^2 - 2|\overline{AB}|\cdot|\overline{AC}|\cos\angle(\overline{AB}, \overline{AC}),$$

т.е. в обычных геометрических обозначениях

$$a^2 = b^2 + c^2 - 2bc\cos\angle A.$$

Полученное соотношение называют *теоремой косинусов*. Если угол A прямой, то $\cos\angle A = 0$ и теорема косинусов превращается в *теорему Пифагора*:

$$a^2 = b^2 + c^2.$$

Условие перпендикулярности

Теорема. Два ненулевых вектора перпендикулярны тогда и только тогда, когда их скалярное произведение равно нулю.

Доказательство. Пусть $\mathbf{a}$ и $\mathbf{b}$ — векторы, φ — угол между ними. Если $\mathbf{a} \perp \mathbf{b}$, то $\cos\varphi = 0$ и $\mathbf{a}\cdot\mathbf{b} = 0$.
Обратно: если $\mathbf{a}\cdot\mathbf{b} = 0$ и при этом $\mathbf{a} \neq 0$, $\mathbf{b} \neq 0$, то $\cos\varphi = 0$ и $\mathbf{a} \perp \mathbf{b}$.
Теорема доказана.

Эта теорема имеет геометрические применения, так как, вычисляя скалярное произведение, легко проверять перпендикулярность векторов, а значит, и прямых.

Пример

1. Рассмотрим четырехугольник $ABCD$ (рис. 98). Пусть K, L, M, N — середины соответствующих сторон. Доказать, что если $|KL| = |MN|$, то диагонали AC и BD перпендикулярны между собой.

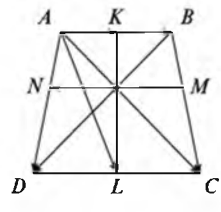


Рис. 98

Доказательство. В векторных обозначениях $\overline{AK} = \frac{1}{2}\overline{AB}$, $\overline{AL} = \frac{1}{2}(\overline{AC} + \overline{AD})$ и поэтому $\overline{KL} = \frac{1}{2}(\overline{AC} + \overline{AD} - \overline{AB}) = \frac{1}{2}(\overline{AC} + \overline{BD})$. Аналогично, $\overline{MN} = \frac{1}{2}(\overline{BD} - \overline{AC})$. Складывая почленно, получаем $\overline{BD} = \overline{KL} + \overline{MN}$.

Аналогично, $\overline{AC} = \overline{KL} - \overline{MN}$. Перемножая, получаем $\overline{BD} \cdot \overline{AC} = (\overline{KL} + \overline{MN})(\overline{KL} - \overline{MN}) = \overline{KL}^2 - \overline{MN}^2 = 0$. Значит, $\overline{AC} \perp \overline{BD}$, что и требовалось доказать.

Контрольные вопросы и задания

1. Что такое скалярное произведение?
2. Приведите пример физической величины, вычисляемой с помощью скалярного произведения.
3. Как выражается скалярное произведение через проекции?
4. Какие вы знаете свойства скалярного произведения?
5. В чем состоит теорема косинусов?
6. Как выразить условие перпендикулярности двух векторов?

§ 10. Координаты вектора

Разложение векторов на составляющие

Построение в плоскости. Пусть в некоторой плоскости даны вектор **a** и две непараллельные прямые. Возьмем произвольную точку O, отложим от нее вектор $\mathbf{a} = \overline{OA}$. Проведем через точки O и A прямые, параллельные двум данным прямым. Получим параллелограмм OBAC. Вектор $\overline{OA}$ можно представить так: $\overline{OA} = \overline{OB} + \overline{OC}$. Полученное представление называется *разложением вектора по двум данным направлениям в плоскости* (рис. 99).

Построение в пространстве. Пусть даны вектор **a** и три прямые, попарно не параллельные и не лежащие все три в одной плоскости. Возьмем произвольную точку O, отложим от нее вектор $\mathbf{a} = \overline{OA}$. Проведем через точку O прямые l_1, l_2, l_3, параллельные данным трем прямым. Построим параллелепипед с ребрами, параллельными l_1, l_2, l_3, и с диагональю *OA*. По правилу параллелепипеда получим разложение $\overline{OA} = \overline{OB} + \overline{OC} + \overline{OD}$, которое называется ***разложением вектора по трем данным направлениям в пространстве*** (рис. 100).

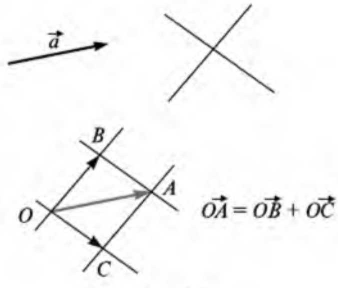


Рис. 99

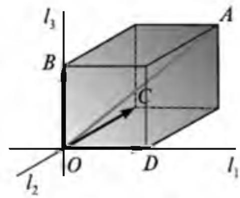


Рис. 100

Примеры

Представим себе, что мы тянем сани за веревку с некоторой силой **F**. Нас может интересовать, какой эффект производит эта сила в двух направлениях: горизонтальном и вертикальном. Для этого надо представить вектор **F** в виде суммы векторов $\mathbf{F}_1$ и $\mathbf{F}_2$, направленных по вертикали и горизонтали (рис. 101).

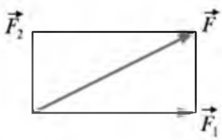


Рис. 101

Рис. 102

Направления, по которым приходится раскладывать вектор, не обязательно совпадают с горизонтальным и вертикальным.

Например, силу тяжести груза (рис. 102) нужно разложить по направлениям удерживающих тросов, чтобы узнать, выдержат ли тросы этот груз.

Размерность

Сравнивая разложение векторов на прямой, в плоскости и в пространстве, можно заметить некоторые общие закономерности. Так, на прямой можно выбрать одно, на плоскости — два, в пространстве — три направления, по которым можно разложить любой вектор (прямой, плоскости, пространства). С другой стороны, если взять на прямой два, на плоскости — три, в пространстве — четыре вектора, то всегда один из них будет раскладываться по направлениям, задаваемым остальными векторами. Последнее свойство прямой, плоскости и пространства мы не доказываем, считая его очевидным. Эти свойства обычно принимают в качестве аксиом, указывающих размерность прямой, плоскости и пространства. Размерность прямой считается равной единице, плоскости — двум, пространства — трем. Это соответствует нашим представлениям о числе «измерений» прямой, плоскости и пространства.

Декартова система координат

Определим координаты точки в пространстве по аналогии с координатами на прямой и на плоскости. Для этого выберем произвольную начальную точку O и проведем через нее три взаимно перпендикулярные оси с общим началом O. Обозначим эти оси Ox, Oy, Oz.

Плоскость, натянутую на Oy и Oz, обозначим yOz; плоскость, натянутую на Ox и Oy, — xOy; плоскость, натянутую на Oz и Ox, — zOx.

Пусть M — произвольная точка. Проведем через нее плоскости, параллельные плоскостям xOy, yOz, zOx. Они вместе с координатными плоскостями (xOy, yOz, zOx) образуют параллелепипед. В нем три ребра OM_x, OM_y, OM_z расположены на осях координат и OM является диагональю. Положение точек на осях Ox, Oy, Oz определяется координатами точек. Обозначим координаты точек M_x, M_y, M_z через x, y, z соответственно и назовем их координатами точки M (рис. 103).

Каждой точке пространства сопоставим набор из трех чисел — координат этой точки. Ясно, что при этом для любой тройки чисел (x; y; z) в пространстве найдется точка с координатами (x; y; z). Таким образом, мы установили взаимно однозначное соответствие между точками пространства и тройками чисел.

В дальнейшем пространство, снабженное декартовой системой координат, мы будем называть *координатным пространством.*

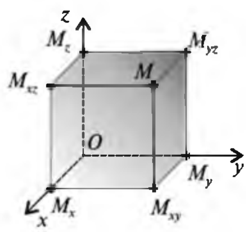


Рис. 103

Нам известна важнейшая математическая идея (связываемая с именами французских математиков XVII в. Декарта и Ферма) задавать точки числами, координатами. Теперь выясним, как, следуя этой идее, задавать векторы. Это расширит наши возможности в действиях с векторами, а также приведет к интересным обобщениям.

Определение координат вектора

Произвольный вектор a отложим от начала координат. Тогда он изобразится направленным отрезком $\overline{OA}$. Координаты его конца, т.е. точки A, называются координатами вектора a (рис. 104).

Это определение имеет смысл как на координатной прямой, координатной плоскости, так и в координатном пространстве.

Итак, чтобы найти координаты вектора, надо отложить его от начала координат и взять координаты конца построенного отрезка.

Координаты вектора обозначаются так же, как и координаты точек, т.е. наборами чисел, записанными в скобках.

С помощью разложения векторов на составляющие можно иначе описать нахождение координат вектора. Координатные оси задают направления (два направления на плоскости, три — в пространстве). Разложим вектор по этим направлениям. Каждая составляющая, расположенная на одной оси, может быть задана одним числом. Вектор задается набором составляющих, а значит, набором чисел, которые и называются *координатами вектора*.

Действия над векторами в координатах

Рассмотрим на плоскости декартову систему координат. Единичные векторы, направленные по осям координат x и y, обозначим соответственно **i** и **j** и назовем *ортами координатных осей* (рис. 105).

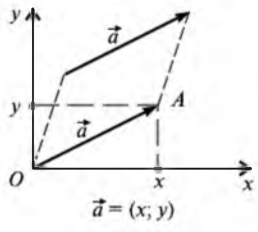


Рис. 104

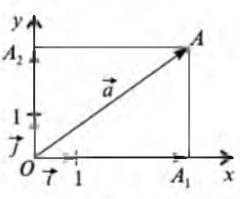


Рис. 105

Возьмем произвольный вектор **a**. Изобразим его направленным отрезком $\overline{OA}$. Разложим $\overline{OA}$ по направлениям координатных осей:

$$\overline{OA} = \overline{OA_1} + \overline{OA_2}.$$

Пусть $(x; y)$ — координаты точки A, т.е. x — координата точки A_1 на оси абсцисс, y — координата точки A_2 на оси ординат. На векторном языке это означает, что $\overline{OA_1} = x\mathbf{i}$ и $\overline{OA_2} = y\mathbf{j}$. Получаем

$$\mathbf{a} = \overline{OA} = \overline{OA_1} + \overline{OA_2} = x\mathbf{i} + y\mathbf{j}.$$

Разложение $\mathbf{a} = x\mathbf{i} + y\mathbf{j}$ называется ***разложением вектора по ортам координатных осей.*** Числа x и y в этом разложении являются координатами точки A, а следовательно, по определению координат вектора, координатами вектора **а**. Итак,

координаты вектора а равны коэффициентам его разложения по ортам координатных осей.

Аналогичное построение делается в пространстве.

Рассмотрим координатное пространство и обозначим **i**, **j**, **k** единичные векторы, направленные по осям координат (орты).

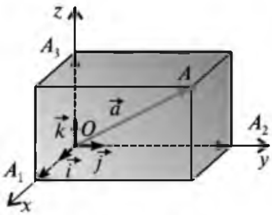


Рис. 106

Возьмем произвольный вектор **а**, отложим его от начала координат: $\mathbf{a} = \overline{OA}$. Если точка A имеет координаты (x; y; z), то вектор **а** представляется в виде $\mathbf{a} = \overline{OA} = \overline{OA_1} + \overline{OA_2} + \overline{OA_3} = x\mathbf{i} + y\mathbf{j} + z\mathbf{k}$ (рис. 106), т.е. коэффициенты разложения вектора по ортам являются координатами этого вектора.

Следующие правила показывают, как выполнять действия с векторами на языке координат.

Правило 1

При умножении вектора на число его координаты умножаются на это число.

Доказательство. Согласно условию, $\mathbf{a} = x\mathbf{i} + y\mathbf{j}$. Следовательно, $\lambda\mathbf{a} = \lambda(x\mathbf{i} + y\mathbf{j})$ и $\lambda\mathbf{a} = (\lambda x)\mathbf{i} + (\lambda y)\mathbf{j}$. Последнее равенство означает, что вектор $\lambda\mathbf{a}$ имеет координаты (λx; λy). На рисунке 107 приведена иллюстрация этого правила.

Правило 2

При сложении векторов их координаты складываются.

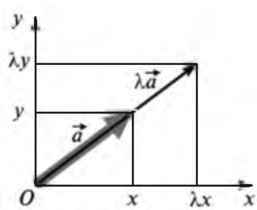


Рис. 107

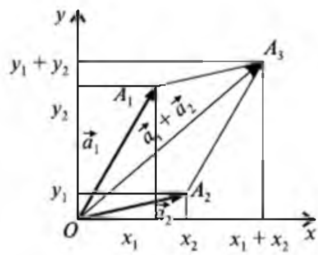


Рис. 108

Доказательство. Согласно условию, $\mathbf{a}_1 = x_1\mathbf{i} + y_1\mathbf{j}$, $\mathbf{a}_2 = x_2\mathbf{i} + y_2\mathbf{j}$. Складывая эти равенства, получаем

$$\mathbf{a}_1 + \mathbf{a}_2 = (x_1 + x_2)\mathbf{i} + (y_1 + y_2)\mathbf{j}.$$

Это равенство означает, что вектор $\mathbf{a}_1 + \mathbf{a}_2$ имеет координаты $(x_1 + x_2;\ y_1 + y_2)$.

Иллюстрацией к этому правилу служит рис. 108. На нем векторы $\mathbf{a}_1$, $\mathbf{a}_2$ показаны радиусами-векторами точек A_1, A_2.

Сумма векторов $\mathbf{a}_1$, $\mathbf{a}_2$ изображается радиусом-вектором точки A_3 — вершины параллелограмма, построенного на OA_1 и OA_2. На рисунке видно, что координаты точки A_3 равны суммам координат точек A_1 и A_2.

Для векторов в пространстве сформулированные правила действий сохраняются.

1. Если вектор $\mathbf{a}$ имеет координаты $(x; y; z)$, то вектор $\lambda\mathbf{a}$, где $\lambda \in R$, имеет координаты $(\lambda x; \lambda y; \lambda z)$.
2. Если вектор $\mathbf{a}_1$ имеет координаты $(x_1; y_1; z_1)$, а вектор $\mathbf{a}_2$ имеет координаты $(x_2; y_2; z_2)$, то вектор $\mathbf{a}_1 + \mathbf{a}_2$ имеет координаты $(x_1 + x_2; y_1 + y_2; z_1 + z_2)$.

Для нахождения координат вектора у нас есть два правила. Согласно первому правилу (принятому за определение), надо вектор отложить от начала координат и взять обычные декартовы координаты точки — конца построенного отрезка. Согласно второму правилу, вектор надо разложить по ортам координатных осей и взять коэффициенты этого разложения. Есть еще третье правило — надо изобразить вектор направленным отрезком и взять разности координат конца и начала.

Теорема. Координаты вектора равны разности координат конца и начала направленного отрезка, изображающего этот вектор (рис. 109).

Доказательство. Пусть вектор $\mathbf{a}$ с координатами $(x; y)$ изображен направленным отрезком $\overline{AB}$. Представим вектор $\mathbf{a}$ как разность радиусов-векторов конца и начала: $\mathbf{a} = \overline{AB} = \overline{OB} - \overline{OA}$. Пусть $(x_A; y_A)$ и $(x_B; y_B)$ — координаты точек A и B. Эти же числа являются координатами векторов OA и OB. Так как при вычитании векторов вычитаются их координаты, то

$$x = x_B - x_A, \quad y = y_B - y_A.$$

Аналогичное правило верно и для векторов в пространстве.

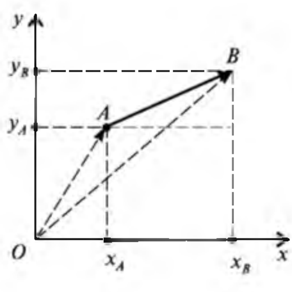


Рис. 109

Выражение скалярного произведения в координатах

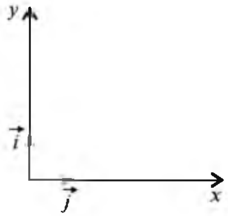


Рис. 110

Введем на плоскости декартову систему координат. Пусть **i** и **j** — орты координатных осей. Так как **i** ⊥ **j**, а длины их равны единице, то $\mathbf{i}^2 = \mathbf{j}^2 = 1$, а $\mathbf{ij} = \mathbf{ji} = 0$ (рис. 110).

Возьмем теперь два вектора $\mathbf{a}_1$ и $\mathbf{a}_2$ на плоскости и разложим их по ортам:

$$\mathbf{a}_1 = x_1\mathbf{i} + y_1\mathbf{j},\ \mathbf{a}_2 = x_2\mathbf{i} + y_2\mathbf{j}.$$

Вычислим скалярное произведение векторов $\mathbf{a}_1$ и $\mathbf{a}_2$. При этом вычислении будем учитывать линейность скалярного произведения, а также правило умножения ортов:

$$\mathbf{a}_1\mathbf{a}_2 = (x_1\mathbf{i} + y_1\mathbf{j})(x_2\mathbf{i} + y_2\mathbf{j}) =$$
$$= (x_1\mathbf{i})(x_2\mathbf{i}) + (x_1\mathbf{i})(y_2\mathbf{j}) + (y_1\mathbf{j})(x_2\mathbf{i}) + (y_1\mathbf{j})(y_2\mathbf{j}) = x_1x_2 + y_1y_2.$$

Сформулируем *правило вычисления скалярного произведения в координатах*:

скалярное произведение двух векторов на плоскости равно сумме произведений их одноименных координат:

$$\mathbf{a}_1\mathbf{a}_2 = x_1x_2 + y_1y_2.$$

Из этой формулы можно вывести ряд следствий.

1. Длина вектора **a** с координатами $(x; y)$:

$$|\mathbf{a}| = \sqrt{\mathbf{aa}} = \sqrt{x^2 + y^2}\ .$$

2. Условие перпендикулярности:

$$\mathbf{a}_1 \perp \mathbf{a}_2 \Leftrightarrow \mathbf{a}_1\mathbf{a}_2 = 0 \Leftrightarrow x_1x_2 + y_1y_2 = 0.$$

3. Формула для угла φ между векторами:

$$\cos\varphi = \frac{\mathbf{a}_1\mathbf{a}_2}{|\mathbf{a}_1|\cdot|\mathbf{a}_2|} = \frac{x_1x_2 + y_1y_2}{\sqrt{x_1^2 + y_1^2}\sqrt{x_2^2 + y_2^2}}\ .$$

Аналогичные формулы верны и в пространстве. Если **i**, **j**, **k** — орты координатных осей в пространстве, то их скалярные квадраты равны 1, а скалярные произведения различных ортов равны нулю:

$$\mathbf{i}^2 = \mathbf{j}^2 = \mathbf{k}^2 = 1,\ \mathbf{ij} = \mathbf{jk} = \mathbf{ki} = 0.$$

Все рассуждения теперь проведем точно так же. Разложим векторы $\mathbf{a}_1$ и $\mathbf{a}_2$ по их ортам:

$$\mathbf{a}_1 = x_1\mathbf{i} + y_1\mathbf{j} + z_1\mathbf{k},\ \ \mathbf{a}_2 = x_2\mathbf{i} + y_2\mathbf{j} + z_2\mathbf{k}.$$

Вычислим $\mathbf{a}_1\mathbf{a}_2$, используя линейность скалярного произведения и правило умножения ортов:

$$\mathbf{a}_1\mathbf{a}_2 = x_1x_2 + y_1y_2 + z_1z_2,$$

т.е. скалярное произведение двух векторов равно сумме произведений их одноименных координат.

Запишем несколько полезных формул.

1. Длина вектора $\mathbf{a}$ с координатами $(x; y; z)$ $|\mathbf{a}|^2 = \mathbf{aa} = x^2 + y^2 + z^2$, откуда

$$|\mathbf{a}| = \sqrt{x^2 + y^2 + z^2}.$$

2. Расстояние между двумя точками в пространстве. Пусть даны две точки: $A_1(x_1; y_1; z_1)$ и $A_2(x_2; y_2; z_2)$. Расстояние между ними можно вычислить как длину вектора $\overline{A_1A_2}$. Координаты вектора $\overline{A_1A_2}$ таковы: $(x_2 - x_1; y_2 - y_1; z_2 - z_1)$. Тогда

$$|A_1A_2| = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2}.$$

3. Условие перпендикулярности двух векторов в пространстве:

$$\mathbf{a}_1 \perp \mathbf{a}_2 \Leftrightarrow \mathbf{a}_1\mathbf{a}_2 = 0 \Leftrightarrow x_1x_2 + y_1y_2 + z_1z_2 = 0.$$

«Ферма... раньше Декарта понял принцип аналитической геометрии.. и установил основной принцип, по которому уравнение первого порядка на плоскости представляет прямую».

Н. Бурбаки

Ферма Пьер (1605—1665) — французский математик, один из создателей аналитической геометрии и дифференциального исчисления. Открыл правило нахождения экстремума с помощью производной. Автор многих теорем теории чисел. Знаменитая теорема Ферма из теории чисел, которую Ферма сформулировал без доказательства, доказана английским математиком Уайлсом только в 1993 г. Формулировка Ферма гласит: «Разделить куб на два других куба и вообще какую-нибудь степень выше второй на две степени с тем же обозначением невозможно, и я нашел воистину замечательное доказательство этого, но поля слишком узки, чтобы вместить его».

Контрольные вопросы и задания

1. Опишите процедуру разложения вектора плоскости по двум направлениям.
2. Сколько направлений нужно выбрать в пространстве, чтобы по ним можно было разложить любой вектор?

3. Как вычисляются декартовы координаты точек?
4. Как определяются координаты вектора?
5. Как связаны координаты вектора с его разложением по координатным осям?
6. Как ведут себя координаты при действиях над векторами?
7. Как вычислить координаты вектора, если известны координаты его начала и конца?
8. Как вычисляется скалярное произведение в координатах?
9. Как вычисляется длина вектора с помощью координат?
10. Как в координатной форме записать условие перпендикулярности двух векторов в пространстве?

§ 11. Применение векторов в механике и геометрии

Закон движения материальной точки

Рассмотрим материальную точку P. Если точка P движется по прямой, то ее положение на прямой задается одним числом. Величины, связанные с движением точки P по прямой, например скорость, сила, ускорение, являются скалярными. Если точка P движется по криволинейной траектории, то все эти величины будут определяться не только числовыми значениями (модулем), но и направлением, т.е. они будут векторами.

Зафиксируем некоторую точку отсчета O и будем положение движущейся точки в момент времени t задавать радиусом-вектором относительно O (рис. 111).

Если в моменты времени t_1, t_2, t_3 точка занимает положения A_1, A_2, A_3, то ее радиусы-векторы $\mathbf{r}(t_1) = \overline{OA_1}$, $\mathbf{r}(t_2) = \overline{OA_2}$, $\mathbf{r}(t_3) = \overline{OA_3}$ будут векторами, зависящими от времени.

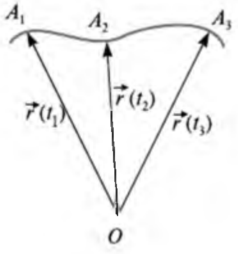


Рис. 111

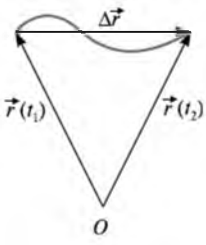


Рис. 112

Итак, мы получили первую векторную величину, связанную с движением точки, — радиус-вектор $\mathbf{r} = \mathbf{r}(t)$, определяющий ее положение относительно некоторой точки отсчета O в момент времени t.

Часто в механике важно знать не положение точки, а ее перемещение за интервал времени $[t_1, t_2]$. Перемещение является вектором и изображается направленным отрезком, начало и конец которого совпадают с положениями точки в моменты t_1 и t_2. Перемещение обозначают $\Delta\mathbf{r}$. Вектор $\Delta\mathbf{r}$ связан с радиусами-векторами, характеризующими положение точки: $\Delta\mathbf{r} = \mathbf{r}(t_2) - \mathbf{r}(t_1)$. Про перемещение можно сказать, что оно является приращением вектора $\mathbf{r}$ за отрезок времени $[t_1, t_2]$ (рис. 112).

В простейшем случае, когда точка движется по прямой, скорость направлена по этой же прямой. В общем случае скорость направлена по касательной к траектории движения (рис. 113).

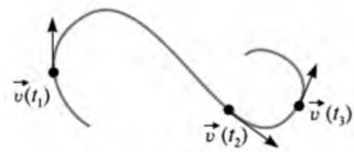


Рис. 113

Положение точки $\mathbf{r}$ является векторной функцией времени: $\mathbf{r} = r(t)$. Эта функция называется *законом движения материальной точки*.

Простейший закон движения — это закон движения с постоянной скоростью. Он задается формулой

$$\mathbf{r} = \mathbf{r}_0 + \mathbf{v}t,$$

где t — время; $\mathbf{v}$ — постоянный вектор скорости; $\mathbf{r}_0$ — положение точки при $t = 0$; $\mathbf{r}$ — переменный вектор; задающий положение точки в момент времени t.

Ясно, что такое движение будет прямолинейным: точка движется по прямой, идущей через конец вектора $\mathbf{r}$ (его начало — фиксированная точка O) в направлении, определяемом вектором $\mathbf{v}$ (рис. 114). Если $\mathbf{v} = 0$, то $\mathbf{r} = \mathbf{r}_0$ и точка стоит на месте.

Квадратичный закон движения $\mathbf{r} = \mathbf{r}_0 + \mathbf{v}_0 t + \dfrac{\mathbf{a}t^2}{2}$, где векторы $\mathbf{r}_0$, $\mathbf{v}_0$ и $\mathbf{a}$ постоянны, задает движение с постоянным ускорением $\mathbf{a}$ и линейно меняющейся скоростью $\mathbf{v} = \mathbf{v}_0 + \mathbf{a}t$ (рис. 115, при $\mathbf{a} = \vec{g}$).

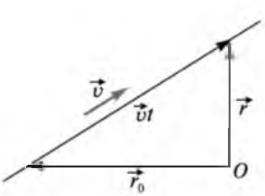


Рис. 114

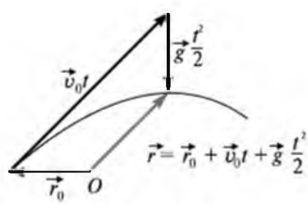


Рис. 115

Изучение векторных законов движения удобно вести с помощью координат. Если материальная точка P движется в координатной плоскости xOy, то ее положение задается ее координатами x и y, которые равны, по определению, координатам ее радиуса-вектора $\mathbf{r}$ (относительно точки O). Таким образом, задать векторную функцию $\mathbf{r} = r(t)$ можно с помощью двух скалярных функций $x = x(t)$ и $y = y(t)$. Переход от векторного задания движения к координатному обычно упрощает решение задач.

Пример

Рассмотрим движение снаряда, начальная скорость $\mathbf{v}_0$ которого была направлена под углом α к горизонту. Выберем в качестве начальной точки O положение снаряда в момент времени $t = 0$, тогда $\mathbf{r} = \mathbf{v}_0 t + \frac{1}{2}\mathbf{g}t^2$. (Здесь мы, конечно, имеем в виду идеальную ситуацию, когда сила тяжести, действующая на снаряд, постоянна и другие силы на снаряд не действуют.)

Выберем оси координат так, как показано на рис. 116. Векторное равенство $\mathbf{r} = \mathbf{v}_0 t + \frac{1}{2}\mathbf{g}t^2$ запишем в координатном виде. Сначала разложим векторы $\mathbf{v}_0$ и $\mathbf{g}$ по горизонтальному и вертикальному направлениям. Проекция вектора $\mathbf{v}_0$ на ось x равна $v_0 \cos\alpha$, а на ось y — $v_0 \sin\alpha$; проекция ускорения на ось x равна нулю, а на ось y равна $-g$. Таким образом,

$$x = t\,v_0 \cos\alpha,\ y = t\,v_0 \sin\alpha - \frac{1}{2}gt^2.$$

где x, y — координаты вектора $\mathbf{r}$. Такая система позволяет решить ряд задач.

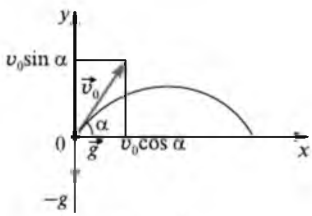


Рис. 116

Задачи

1. Найти время движения снаряда.

Снаряд касается земли в момент $t > 0$, когда $y = 0$. Приравниваем к нулю выражение для y:

$$t\,v_0 \sin\alpha - \frac{1}{2}gt^2 = 0.$$

Решая это квадратное уравнение относительно t, получаем $t_1 = 0$, $t_2 = \frac{2v_0 \sin\alpha}{g}$. Ясно, что t_1 соответствует моменту вылета снаряда, а t_2 — моменту его падения.

2. Найти расстояние по горизонтали, которое пролетит снаряд. Известно время движения снаряда. Подставляя его в формулу для x, получаем $x = \dfrac{2v_0^2 \sin\alpha\cos\alpha}{g} = \dfrac{2v_0^2 \sin\alpha\cos\alpha}{g}$.

Найдите самостоятельно, на какую максимальную высоту поднимается снаряд.

Векторное задание прямой

Направляющий вектор прямой

Направляющим вектором прямой называется любой ненулевой вектор, который лежит на этой прямой.

Рассмотрим в пространстве прямую l. Выберем на ней две точки: A и B. Вектор $\overrightarrow{AB}$ является направляющим вектором прямой l. Направляющий вектор прямой определен неоднозначно. Любой ненулевой вектор, коллинеарный $\overrightarrow{AB}$, также будет направляющим вектором прямой l.

На рисунке 117 изображены различные направленные отрезки, которые можно принять за направляющий вектор прямой l. Обозначим направляющий вектор прямой одной буквой, например $\mathbf{v}$ (рис. 118). Такое обозначение принято в курсе физики: если точка равномерно движется по прямой l, то ее скорость $\mathbf{v}$ может быть взята за направляющий вектор этой прямой. Отметим, что если $\mathbf{v}_1$ и $\mathbf{v}_2$ — направляющие векторы одной и той же прямой l, то они коллинеарны, т.е. найдется такое число $t \neq 0$, что $\mathbf{v}_2 = t\mathbf{v}_1$. Это означает, что если две точки движутся по одной и той же прямой или по параллельным прямым, то их скорости являются коллинеарными векторами.

Чтобы однозначно определить прямую, нужно кроме ее направляющего вектора $\mathbf{v}$ задать положение какой-либо ее точки. Пусть в пространстве зафиксирована начальная точка O, дана точка P_0 своим радиусом-вектором $\mathbf{r}_0$ и дан вектор $\mathbf{v} \neq 0$. Проведем через точку P_0 прямую l с направляющим вектором $\mathbf{v}$.

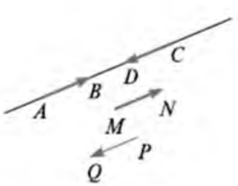


Рис. 117

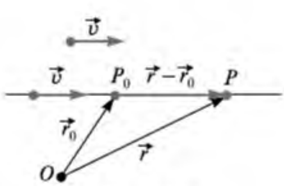


Рис. 118

Пусть $\mathbf{r}$ — радиус-вектор ее произвольной точки P. Тогда вектор $\overline{P_0P}$, как один из направляющих векторов прямой l, коллинеарен вектору $\mathbf{v}$, т.е. найдется такое число t, что $\overline{P_0P} = t\mathbf{v}$. Записывая вектор $\overline{P_0P}$ как разность векторов $\mathbf{r}$ и $\mathbf{r}_0$, получаем: $\mathbf{r} - \mathbf{r}_0 = t\mathbf{v}$ или $\mathbf{r} = \mathbf{r}_0 + t\mathbf{v}$.

Обратно: если радиус-вектор $\mathbf{r}$ какой-либо точки P удовлетворяет такому уравнению при некотором $t \in R$, то вектор $\overline{P_0P}$ коллинеарен $\mathbf{v}$ и точка P лежит на прямой l. Поэтому это уравнение называют *уравнением прямой* (в векторной форме). Оно описывает траекторию точки, движущейся с постоянной скоростью $\mathbf{v}$ и имеющей положение P_0 в момент времени $t = 0$.

Векторное уравнение прямой $\mathbf{r} = \mathbf{r}_0 + t\mathbf{v}$, или $\mathbf{r} - \mathbf{r}_0 = t\mathbf{v}$, можно переписать в координатной форме. Рассмотрим для простоты прямую, лежащую в плоскости с декартовыми координатами (x, y). Пусть (x, y) — координаты переменного вектора $\mathbf{r}$, (x_0, y_0) — координаты точки P_0; (m, n) — координаты направляющего вектора $\mathbf{v}$. Векторное равенство означает равенство соответствующих координат:

$$\begin{cases} x - x_0 = tm, \\ y - y_0 = tn. \end{cases}$$

Мы воспользовались правилом действий, над векторами в координатах. От системы уравнений можно перейти к одному уравнению, связывающему x и y, если исключить параметр t. Для этого надо умножить первое уравнение на n и вычесть из него второе, умноженное на m. Получится уравнение прямой на плоскости в координатной форме:

$$n(x - x_0) - m(y - y_0) = 0.$$

Если раскрыть скобки и обозначить $n = A$, $-m = B$, $-nx_0 + my_0 = C$, то мы получим уравнение прямой $Ax + By + C = 0$.

Пример

Пусть прямая l на координатной плоскости (x, y) проходит через точки $A(3; -2)$ и $B(4; 1)$. Построим точки A и B и проведем через них прямую l. В качестве направляющего вектора прямой l можно взять вектор $\overline{AB}$. Он имеет координаты $(1; 3)$, которые находятся как разности координат конца и начала. За направляющий вектор прямой l можно взять и другие векторы — $\overline{BA}$, $\overline{AC}$ (точка C — середина отрезка AB), $\overline{OD}$ (точка D выбрана так, что $OD \parallel AB$).

В качестве начальной точки можно взять точку A. Тогда векторное уравнение прямой l можно записать в виде $\mathbf{r} = \mathbf{r}_0 + t\mathbf{v}$, где

$\mathbf{r}(x; y)$ — радиус-вектор переменной точки прямой l, $\mathbf{r}_0(3; -2) = \overrightarrow{OA}$, $\mathbf{v}(1; 3) = \overrightarrow{AB}$. В координатах уравнение прямой l имеет вид

$$\begin{cases} x = 3 + t, \\ y = -2 + 3t. \end{cases}$$

Умножая первое уравнение на 3 и вычитая из него второе, получаем уравнение $3x - y = 11$, или $3x - y - 11 = 0$. Это же уравнение можно записать в виде $y = 3x - 11$, откуда следует, что угловой коэффициент прямой l равен 3.

Параллельность и перпендикулярность прямых

Если две прямые параллельны, то их направляющие векторы коллинеарны.

Коллинеарность векторов геометрически означает, что изображающие их направленные отрезки лежат на параллельных прямых. Можно сказать, что направляющий вектор прямой l является одновременно направляющим вектором любой прямой, параллельной l.

Запишем условия параллельности прямых l_1 и l_2 с направляющими векторами $\mathbf{v}_1$ и $\mathbf{v}_2$ (рис. 119):

$$l_1 \parallel l_2 \Leftrightarrow \mathbf{v}_1 \parallel \mathbf{v}_2 \Leftrightarrow \mathbf{v}_1 = \lambda \mathbf{v}_2.$$

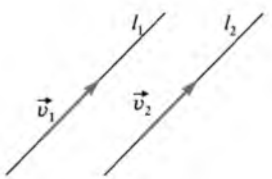


Рис. 119

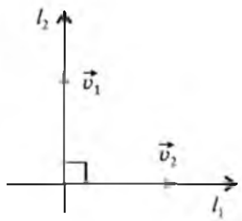


Рис. 120

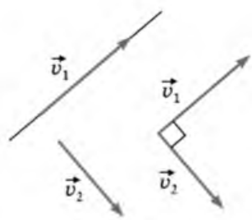


Рис. 121

Рассмотрим теперь перпендикулярные прямые. Возьмем две пересекающиеся прямые l_1 и l_2 с направляющими векторами $\mathbf{v}_1$ и $\mathbf{v}_2$. Прямая l_1 перпендикулярна прямой l_2 тогда и только тогда, когда векторы $\mathbf{v}_1$ и $\mathbf{v}_2$ ортогональны (рис. 120).

Не обязательно брать только пересекающиеся прямые. Две прямые в пространстве называются ***перпендикулярными***, если ортогональны их направляющие векторы (рис. 121).

Запишем условие перпендикулярности прямых l_1 и l_2 с направляющими векторами $\mathbf{v}_1$ и $\mathbf{v}_2$:

$$l_1 \perp l_2 \Leftrightarrow \mathbf{v}_1 \perp \mathbf{v}_2 \Leftrightarrow \mathbf{v}_1 \cdot \mathbf{v}_2 = 0.$$

Примеры

1. Найдите на рис. 122 координаты точек C и D, вычислите координаты векторов $\overrightarrow{BA}$, $\overrightarrow{AC}$ и $\overrightarrow{OD}$ и убедитесь в том, что они коллинеарны вектору $\overrightarrow{AB}$.
2. В кубе $ABCDA'B'C'D'$ (рис. 123) ребро AB перпендикулярно восьми ребрам: AA', BB', CC', DD', AD, $A'D'$, $B'C'$ и BC. Проверим, что диагонали AC и $B'D'$ тоже перпендикулярны. Докажем, что $\overrightarrow{AC} \cdot \overrightarrow{B'D'} = 0$.

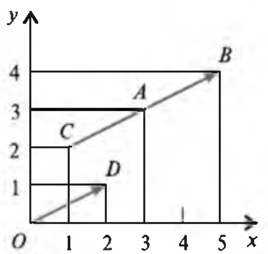


Рис. 122

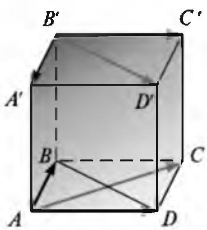


Рис. 123

Сначала заменим вектор $\overrightarrow{B'D'}$ на равный ему вектор $\overrightarrow{BD}$. Затем разложим: $\overrightarrow{AC} = \overrightarrow{AD} + \overrightarrow{AB}$, $\overrightarrow{BD} = \overrightarrow{AD} - \overrightarrow{AB}$. Получим $\overrightarrow{AC} \times \overrightarrow{B'D'} = \overrightarrow{AC} \cdot \overrightarrow{BD} = (\overrightarrow{AD} + \overrightarrow{AB}) \cdot (\overrightarrow{AD} - \overrightarrow{AB}) = \overrightarrow{AD}^2 - \overrightarrow{AB}^2 = 0$, так как длины векторов $\overrightarrow{AD}$ и $\overrightarrow{AB}$ равны как длины сторон квадрата.

3. На следящем устройстве отмечаются координаты двух самолетов, летящих по прямолинейным траекториям. Первый самолет находился при двух наблюдениях в точках $A_1(3;\ 17;\ 10)$ и $A_2(8;\ 11;\ 11)$, второй — в точках $B_1(2;\ 29;\ 11)$ и $B_2(7;\ 33;\ 10)$. Докажем, что их траектории перпендикулярны.

Векторы $\overrightarrow{A_1A_2}$ и $\overrightarrow{B_1B_2}$ можно принять за направляющие векторы траекторий. Вычислим координаты этих векторов $\overrightarrow{A_1A_2}$ $(5;\ -6;\ 1)$, $\overrightarrow{B_1B_2}$ $(5;\ 4;\ -1)$ и составим скалярное произведение: $\overrightarrow{A_1A_2} \cdot \overrightarrow{B_1B_2} = 5 \cdot 5 + (-6) \cdot 4 + 1 \cdot (-1) = 0$, т.е. $\overrightarrow{A_1A_2} \perp \overrightarrow{B_1B_2}$.

Ответьте на трудный вопрос: пересекаются ли траектории самолетов?

Угол между прямыми

Углом между двумя прямыми в пространстве называется угол между их направляющими векторами (рис. 124).

Направляющий вектор прямой определен неоднозначно. Как сказывается эта неоднозначность на определении угла между прямыми? Заменим вектор $\mathbf{v}_1$ на коллинеарный ему вектор $\mathbf{v}'_1$. Пусть $\mathbf{v}'_1 = \lambda\mathbf{v}_1$. Если $\lambda > 0$, то углы между векторами ($\mathbf{v}_1$; $\mathbf{v}_2$) и ($\lambda\mathbf{v}_1$; $\mathbf{v}_2$) равны (рис. 125). Если $\lambda < 0$, то направление вектора изменится на противоположное и угол также изменится (рис. 126).

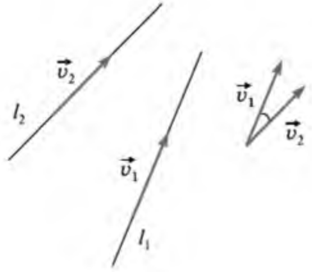


Рис. 124

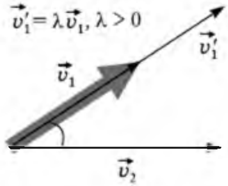


Рис. 125

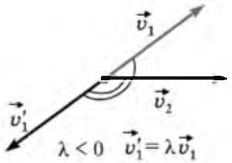


Рис. 126

Таким образом, направляющие векторы двух прямых при различном выборе этих векторов могут образовывать два разных угла, дающих в сумме развернутый угол. Если нужно уточнить, какой из углов имеется в виду, то часто указывают, например, «острый угол между двумя прямыми» или «тупой угол».

Углы между прямыми можно вычислить с помощью скалярного произведения, так как формула для угла между векторами известна. Пусть прямые l_1 и l_2 имеют направляющие векторы $\mathbf{v}_1$ и $\mathbf{v}_2$. Косинус угла между прямыми l_1 и l_2 вычисляется по формуле

$$\cos\angle(\mathbf{v}_1, \mathbf{v}_2) = \frac{\mathbf{v}_1 \cdot \mathbf{v}_2}{|\mathbf{v}_1||\mathbf{v}_2|}.$$

Заметим, что косинусы двух различных углов между прямыми имеют одинаковый модуль и разные знаки, так как $\cos(\pi - \alpha) = -\cos\alpha$. По знаку косинуса можно узнать, острый или тупой угол образуют выбранные направляющие векторы прямых.

Векторное задание плоскости

Нормаль к плоскости

Прямая называется перпендикулярной плоскости, если она перпендикулярна любой прямой на этой плоскости (рис. 127).

Нормалью к плоскости называется ненулевой вектор, который лежит на прямой, перпендикулярной данной плоскости.

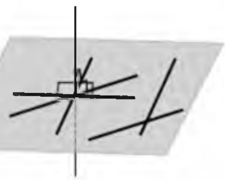


Рис. 127

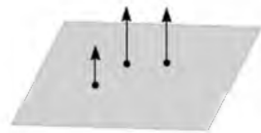


Рис. 128

Нормаль к плоскости определена неоднозначно (рис. 128). Любой ненулевой вектор, коллинеарный нормали, также является нормалью к этой плоскости. Обозначим **n** нормаль к плоскости. Для однозначного задания плоскости кроме нормали достаточно задать положение какой-либо ее точки.

Пусть α — некоторая плоскость, **n** — ее нормаль, $\mathbf{r}_0$ — радиус-вектор какой-либо точки P_0 плоскости α, **r** — радиус-вектор ее произвольной точки P. Так как вектор $\overline{P_0P}$ лежит в плоскости α, то он ортогонален нормали n, т.е. $\overline{P_0P} \cdot \mathbf{n} = 0$ (рис. 129).

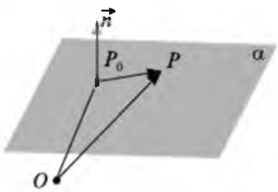


Рис. 129

Обратно: если для некоторой точки P выполняется условие $\overline{P_0P} \times \times \mathbf{n} = 0$, т.е. $\overline{P_0P} \perp \mathbf{n}$, то точка P лежит в плоскости α. Таким образом, условие того, что точка P лежит в плоскости α, имеет вид

$$P \in \alpha \Leftrightarrow \overline{P_0P} \cdot \mathbf{n} = 0.$$

Это условие можно переписать в координатах, зная координаты точки $P_0(x_0;\ y_0;\ z_0)$ и вектора нормали $\mathbf{n}(A;\ B;\ C)$: $(x - x_0) \times \times A + (y - y_0)\ B + (z - z_0)C = 0$. Если раскрыть скобки и обозначить $-Ax_0 - By_0 - Cz_0$ через D, то получим уравнение плоскости в координатах:

$$Ax + By + Cz + D = 0.$$

Если обозначить радиусы-векторы точек P_0 и P через $\mathbf{r}_0$ и $\mathbf{r}$, то векторное уравнение плоскости таково:

$$(\mathbf{r} - \mathbf{r}_0) \cdot \mathbf{n} = 0.$$

Полезно запомнить, что при задании уравнения плоскости в виде $Ax + By + Cz + D = 0$ коэффициенты $(A;\ B;\ C)$ являются координатами вектора нормали к плоскости.

Пример

Прямая l проходит через точки $P_1(1;\ 2;\ -5)$ и $P_2(6;\ -2;\ -3)$. Рассмотрим различные плоскости, перпендикулярные прямой l.

В качестве нормали любой из них можно взять вектор $\overline{P_1P_2}$, т.е. направляющий вектор прямой l. Вектор $\overline{P_1P_2}$ имеет координаты (5; –4; 2). Напишем координатное уравнение плоскости, проходящей через точку P_1 и перпендикулярной прямой l:

$$5(x-1)-4(y-2)+2(z+5)=0,$$

или

$$5x-4y+2z+13=0.$$

Уравнение плоскости, проходящей через точку P_2 и перпендикулярной прямой l, имеет вид

$$5(x-6)-4(y+2)+2(z+3)=0,$$

или

$$5x-4y+2z-32=0.$$

Контрольные вопросы и задания

1. Какой формулой описывается закон движения материальной точки с постоянной скоростью?
2. Какой будет траектория движения материальной точки, если скорость движения постоянна?
3. При каких условиях закон движения материальной точки будет квадратичным?
4. Что такое направляющий вектор прямой?
5. Как записать уравнение прямой в векторной форме?
6. Как записать векторное уравнение прямой в координатной форме?
7. Как проверить, что две прямые, заданные в векторной форме: а) параллельны; б) перпендикулярны?
8. Как определить угол между двумя прямыми?
9. Как записать векторное уравнение плоскости?
10. Как записать уравнение плоскости в координатах? Каков геометрический смысл коэффициентов?

Заключительная беседа

Линейные векторные пространства

Новые примеры векторных величин. До сих пор мы рассматривали вектор как направленный отрезок, т.е. задавали его длиной и направлением. Введение координат позволяет задавать векторы наборами чисел.

Так, вектор плоскости определяется парой чисел, вектор пространства — тройкой чисел. Для того чтобы не противопоставлять скалярные величины векторным, можно считать, что скалярные величины изображаются на прямой и задаются модулем (абсолютным значением, длиной) и направлением (знаком). Задание векторов наборами чисел позволяет рассматривать новые векторные величины.

Допустим, цех выпускает изделия четырех наименований. Выпуск продукции за месяц можно охарактеризовать строчкой из четырех чисел (a_1, a_2, a_3, a_4), где a_1 — количество выпускаемых изделий первого вида, a_2, a_3, a_4 — количество изделий остальных видов. Полностью определить выпуск продукции с помощью одного числа невозможно. Строчка (a_1, a_2, a_3, a_4), характеризующая выпуск продукции по ассортименту, является векторной величиной.

Рассмотрим множество квадратных трехчленов $ax^2 + bx + c$, где a, b, c — произвольные действительные числа (мы допускаем обращение в нуль любого из коэффициентов a, b, c). Например, тройка чисел $(1, 0, 0)$ задает трехчлен вида $x^2 + 0x + 0 = x^2$, а тройка чисел $(2, -1, -1)$ — трехчлен $2x^2 - x - 1$. Квадратный трехчлен мы можем тоже рассматривать как вектор.

Как же выделить какие-то общие свойства векторных величин? Важным свойством, объединяющим все векторные величины, является возможность совершать с ними две операции: сложение и умножение на число.

Сложение параллельных переносов и умножение их на число известно из геометрии, сложение сил — из механики. Если рассматривать строки из четырех чисел (a_1, a_2, a_3, a_4), задающих выпуск фиксированных четырех видов продукции, то естественно складывать их следующим образом:

$$(a_1, a_2, a_3, a_4) + (b_1, b_2, b_3, b_4) = (a_1 + b_1, a_2 + b_2, a_3 + b_3, a_4 + b_4).$$

Это делается при суммировании выпуска продукции за несколько месяцев или определении выпуска продукции несколькими предприятиями, имеющими одинаковый ассортимент изделий.

Аналогично складываются и умножаются на число квадратные трехчлены.

Обратим внимание на важную особенность разобранных примеров. Мы говорили о квадратных трехчленах, но при этом не фиксировали внимание на каком-то одном из них, а брали сразу множество всех квадратных трехчленов. При изучении параллельных переносов полезно рассматривать все параллельные переносы, лишь тогда можно складывать их и умножать на число.

Таким образом, мы каждый раз имеем дело с множеством значений векторной величины, причем эти значения можно складывать и умножать на число. Во всех примерах эти две операции удовлетворяют всем законам векторной алгебры. Эти правила мы уже фактически использовали в доказательствах, но не обращали на это внимания.

В математике множество объектов, которые можно складывать и умножать на число, называют *векторным пространством*, если для этих операций выполнены законы векторной алгебры.

Размерность. Мы начали изучение векторов с того, что указали величины, которые нельзя задавать одним числом. Оказалось возможным задавать эти величины несколькими числами, их координатами.

Так, силу, действующую на точку в пространстве, можно задавать тремя числами — проекциями силы на оси координат. Выпуск продукции цехом, изготовляющим четыре вида изделий, задается четырьмя числами — количеством выпускаемых изделий по каждому виду отдельно. Квадратный трехчлен $ax^2 + bx + c$ задается тремя числами — коэффициентами a, b, c. Таким образом, векторные величины различаются тем, сколько чисел требуется вычислить для их задания (конечно, речь идет о наименьшем количестве чисел, нужных для вычислений векторной величины).

Если для задания векторной величины нужно задать n чисел, то говорят, что она является n-мерной векторной величиной, а про векторное пространство, образованное значениями этой величины, говорят, что оно имеет размерность n, или n-мерно.

Так, силы, действующие в пространстве, описываются с помощью векторов с тремя координатами, размерность пространства параллельных переносов в плоскости равна двум, квадратные трехчлены заполняют трехмерное пространство, а выпуск продукции четырех видов изображается элементом такого пространства, для которого $n = 4$. Скалярные величины задаются одним числом. Их можно рассматривать как одномерные векторные величины.

Примером n-мерного пространства при любом n является множество строчек длины n: $\mathbb{R}^n = \{(a_1, a_2, a_3, a_4, \ldots, a_n)\}$.

Сложение строчек и умножение их на число производится по следующим правилам:

$$(a_1, a_2, a_3, a_4, \ldots, a_n) + (b_1, b_2, b_3, b_4, \ldots, b_n) =$$
$$= (a_1 + b_1, a_2 + b_2, a_3 + b_3, a_4 + b_4, \ldots, a_n + b_n),$$
$$\lambda(a_1, a_2, \ldots, a_n) = (\lambda a_1, \lambda a_2, \ldots, \lambda a_n).$$

Размерность векторной величины является ее основной характеристикой. Все векторные величины одной и той же размерности похожи друг на друга, например, тем, что каждую из них можно задавать строчкой из такого же количества чисел. Это сходство проявляется и в геометрическом изображении векторов. Все двумерные векторные величины можно изобразить на плоскости; все трехмерные — в пространстве. Прибавим к этому, что все одномерные векторные величины — скаляры — изображаются на числовой прямой.

Многие теоретические и прикладные дисциплины (физика, экономика, радиотехника и др.) используют n-мерные векторные пространства с $n > 3$.

Контрольные вопросы и задания

1. Каковы общие свойства векторных величин?
2. Сформулируйте свойства сложения векторов и умножения вектора на число.
3. Приведите примеры величин, задающихся наборами чисел.
4. Приведите пример четырехмерного пространства.

ГЛАВА 3

ТРИГОНОМЕТРИЯ

§ 12. Вращательное движение

Периодические процессы

В этой главе будет изучаться новый класс функций — *тригонометрические функции*. Они служат, прежде всего, для описания разнообразных периодических процессов. С периодически повторяющимися ситуациями человек сталкивается повсюду. Его жизнь сопровождают различные астрономические явления — восход и заход Солнца, изменение фаз Луны, чередование времен года, положение звезд на небе, затмения и движение планет. Человек давно заметил, что все эти явления возобновляются периодически. Жизнь на Земле тесно связана с ними, и поэтому не удивительно, что астрономические наблюдения явились источником многих математических открытий.

Биение сердца, цикл в жизнедеятельности организма, вращение колеса, морские приливы и отливы, заполненность городского транспорта, эпидемии гриппа. В этих многообразных примерах можно найти общее: эти процессы периодичны, состояния участвующих в них объектов повторяются.

Мы часто слышим по радио сообщения об очередном запуске искусственного спутника Земли. Обычно в сообщении указываются наименьшее и наибольшее расстояния спутника от поверхности Земли и период его обращения. Если период обращения спутника составляет 92 мин, то мы понимаем, что его положение относительно Земли через каждые 92 мин будет одинаковым. Так мы приходим к понятию периодической функции как функции, обладающей периодом, т.е. таким числом T, что значения функции при значениях аргумента, отличающихся на T, $2T$, $3T$ и т.д., будут одинаковыми.

Астрономия, которая дает нам наиболее наглядное представление о периодических процессах, определяет положение объектов на небесной сфере с помощью углов. Можно сказать так: в качестве аргумента периодических функций очень часто выступает угол. Поэтому сначала мы обсудим вопрос об измерении углов.

Углы и их измерение

Геометрически угол — это часть плоскости, ограниченная двумя лучами, выходящими из этой точки. Эти лучи называются сторонами угла, а точка, из которой они выходят, — вершиной (рис. 130).

Рис. 130

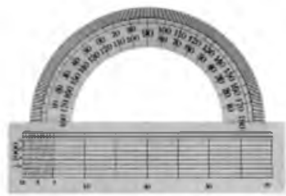


Рис. 131

Как измеряют углы? В качестве единицы геометрических углов принят градус, т.е. $\frac{1}{360}$ полного угла. Конкретные углы удобно измерять в градусах с помощью транспортира (рис. 131).

Во многих оптических приборах также используют градусную меру угла. Углы, получающиеся при непрерывном вращении, удобно измерять не в градусах, а с помощью таких чисел, которые отражали бы процесс построения угла, т.е. вращение.

Представим себе, что зафиксирована не только вершина угла, но и один из образующих его лучей (рис. 132). Другой луч вращается вокруг вершины. Получающиеся углы будут зависеть от угловой скорости вращения и времени. Можно считать, что вращение происходит равномерно (с постоянной угловой скоростью). Тогда угол поворота определяется длиной дуги, по которой проходит какая-либо фиксированная точка подвижного луча.

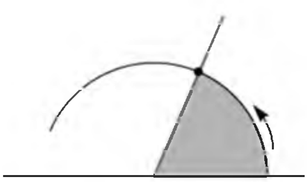


Рис. 132

Если расстояние от точки до вершины равно R, то при вращении точка движется по окружности радиуса R. Отношение пройденного пути к радиусу R не зависит от радиуса и может быть взято за меру угла. Численно она равна пути, пройденному точкой по окружности единичного радиуса. Итак,

пусть угол получен вращением подвижного луча от некоторого начального положения. Его значение численно равно пути, который пройдет точка этого луча, находящаяся на единичном расстоянии от вершины.

Чему равен полный угол? Ясно, что значение полного угла равно длине единичной окружности, т.е. числу 2π (где π — отношение длины окружности к ее диаметру). Число π было известно людям с глубокой древности и с довольно большой точностью. Первые десятичные знаки этого числа таковы: $\pi = 3{,}14159265358...$

Итак, полный угол равен 2π, развернутый — π, а прямой — $\frac{\pi}{2}$. Угол величины 1 называют *радианом* (рад), а введенную меру угла — *радианной*.

Радианную меру легко связать с градусной. Достаточно сравнить меры одного и того же угла, например прямого: $\frac{\pi}{2}$ рад = 90°, откуда

$$1^\circ = \frac{\pi}{180^\circ} \approx 0{,}017 \text{ рад.}$$

Обратно: можно выразить 1 рад в градусной мере, т.е.

$$1 \text{ рад} = \frac{2}{\pi} \cdot 90^\circ = \frac{180}{\pi} \cdot 1^\circ = ... = 57{,}296^\circ.$$

В географии, астрономии и других прикладных науках используют доли градуса — минуту (′) и секунду (″). Минута — это 1/60°, а секунда — 1/60′. Запишем соотношения между различными единицами углов:

$$1^\circ = 0{,}017 \text{ рад};$$

$$1' = \left(\frac{1}{60}\right)' = 0{,}0003 \text{ рад};$$

$$1'' = \left(\frac{1}{60}\right)' = \left(\frac{1}{3600}\right)^\circ = 0{,}000005 \text{ рад};$$

$$1 \text{ рад} = 57{,}296^\circ = 57^\circ\, 17'\, 45''.$$

Градусный и радианный способы измерения углов равноправны и используются достаточно широко. Так, широту и долготу точки на земном шаре обычно задают в градусах и минутах. При проведении приближенных вычислений обычно заменяют $\sin\alpha$ на α (при малых углах α), но это можно делать только тогда, когда угол выражен в радианах.

Заметим еще, что обозначение градуса (минуты, секунды) записывают, а обозначение радиана, как правило, пропускают (так как с физической точки зрения угол — безразмерная величина, например, $\alpha = 0{,}23$; $\alpha = 3{,}14$; $\alpha = 0{,}01$).

Угол π часто используется как самостоятельная единица масштаба, с помощью которой измеряют другие углы — прямой угол равен $\frac{\pi}{2}$, угол в равностороннем треугольнике — $\frac{\pi}{3}$. Часто встречаются записи меры углов в виде $\frac{5}{6}\pi$, $\frac{2}{7}\pi$, $\frac{\pi}{4}$ и т.д. Обычная единица масштаба радиан соответствует некоторому углу, чуть меньшему, чем $\frac{\pi}{3}$, так как $\frac{\pi}{3} = 1{,}047$.

Так как на практике приходится иметь дело как с градусной, так и с радианной мерой, то на микрокалькуляторе есть возможность выбирать единицы измерения угла, т.е. умеет переводить градусы в радианы и обратно.

Подведем итоги.

Угол мы можем получить вращением подвижного луча. Его радианная мера численно равна пути, который проходит точка этого луча, отстоящая от вершины на расстоянии 1.

Движение точки по окружности во многом аналогично движению точки по прямой. Чтобы определить положение точки на прямой, недостаточно знать путь, пройденный ею от начальной точки, нужно указать еще направление движения. Обычно на прямой фиксируют положительное направление, а положение точки определяют одним числом, которое может быть не только положительным (как путь), но и отрицательным.

Аналогично поступают и с вращательным движением. В качестве положительного направления движения по окружности выбирается движение против хода часовой стрелки. Угол задают произвольным числом t. Для построения угла t на единичной окружности от неподвижной точки откладывают путь, равный $|t|$, в направлении, определяемом знаком числа t. Таким образом, для произвольного числа t мы построили угол t, определяемый двумя лучами — неподвижным и тем, который проходит через построенную точку (рис. 133).

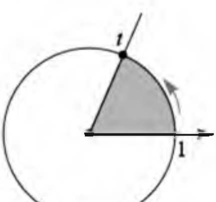


Рис. 133

При таком обобщении понятия угла постепенно отходят от его геометрического образа как части плоскости, лежащей между двумя лучами. Фактически слово «угол» становится для нас синонимом слова «число». Угол t (т.е. произвольное число t) будет выступать у нас в качестве аргумента тригонометрических функций. Изображать угол t нам будет удобно не в виде пары лучей, а в виде точки единичной окружности. Для этого мы подробно рассмотрим вращательное движение.

Вращательное движение и его свойства

Представим себе маленький шарик, который равномерно движется по окружности в положительном направлении. Будем считать, что в момент времени $t = 0$ шарик находился в положении A и за время $t = 1$ прошел по окружности расстояние, равное 1. Тогда за время, равное π, он пройдет расстояние, равное π, т.е. половину окружности.

Обозначим P_t точку на окружности, в которой шарик находится в момент времени t. Для того чтобы найти положение этой точки, надо отложить от точки $P_0 = A$ по окружности путь длины $|t|$ в положительном направлении, если $t > 0$, и в отрицательном направлении (т.е. по ходу часовой стрелки), если $t < 0$.

Примеры

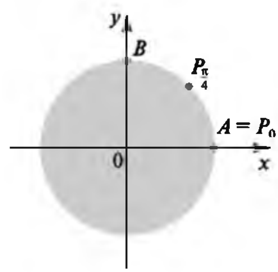


Рис. 134

1. Пусть $t = \dfrac{\pi}{4}$. Отложим от точки P_0 в положительном направлении путь длины $\dfrac{\pi}{4}$. Так как длина всей окружности равна 2π, то точка $P_{\frac{\pi}{4}}$ является серединой дуги AB (рис. 134).
2. Пусть $t = \dfrac{9\pi}{4}$. Отложим от точки P_0 путь длины $\dfrac{9\pi}{4}$. Заметим, что $\dfrac{9\pi}{4} = 2\pi + \dfrac{\pi}{4}$. Пройдя путь 2π, точка опять попадает в A. Пройдя оставшийся путь, она попадает в середину дуги AB. Таким образом, точка $P_{\frac{9\pi}{4}}$ совпадает с точкой $P_{\frac{\pi}{4}}$.
3. Найдем теперь точку $P_{-\frac{\pi}{3}}$. Для этого точке необходимо пройти в отрицательном направлении путь $\dfrac{\pi}{3}$ (рис. 135). Оси координат делят плоскость на четыре четверти. В зависимости от того, в ка-

кую часть плоскости попадает точка P_t, говорят о том, в какую четверть попадает угол t. При этом полезно помнить, что 1 рад чуть меньше 60°, т.е. трети развернутого угла.

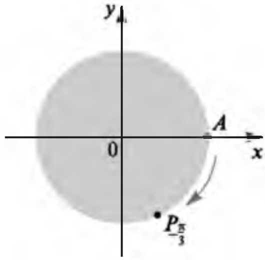


Рис. 135

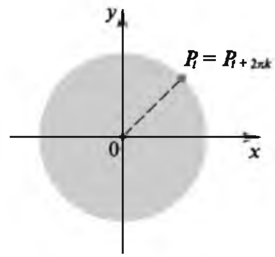


Рис. 136

Свойства вращательного движения

Свойство 1

Для всякого целого числа k точка P_t совпадает с точкой $P_{t+2\pi k}$ (если моменты времени отличаются на число, кратное числу 2π, то точка в эти моменты времени занимает одно и то же положение) (рис. 136).

Это свойство выражает периодичность вращательного движения.

Действительно, в разные моменты времени точка может занимать на окружности одно и то же положение. Так, в примере 2 мы убедились, что $P_{\frac{\pi}{4}} = P_{\frac{9\pi}{4}}$. Это произошло из-за того, что числа $\frac{9\pi}{4}$ и $\frac{\pi}{4}$ отличаются друг от друга на 2π, а за время $t = 2\pi$ точка проходит всю окружность и возвращается в исходное положение. Аналогично, если разность $t_1 - t_2$ равна 4π, 6π, 8π или -2π, -4π и т.д., то $P_{t_1} = P_{t_2}$, т.е. в моменты времени t_1, $t_1 \pm 2\pi$, $t_1 \pm 4\pi$, ... точка занимает одно и то же положение на окружности.

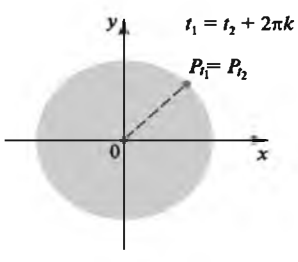


Рис. 137

Свойство 2

Если $P_{t_1} = P_{t_2}$, то найдется такое целое число k, что $t_1 = t_2 + 2\pi k$ (рис. 137).

Действительно, если в два каких-то момента времени положения точки совпали, то между этими двумя моментами точка прошла целое число раз всю окружность.

Свойство 3

Для всякого значения t точки P_t и $P_{t+\pi}$ диаметрально противоположны (рис. 138).

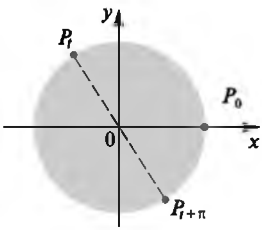


Рис. 138

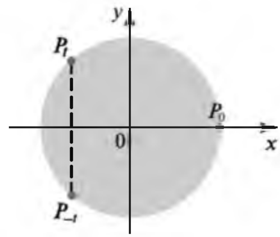


Рис. 139

Действительно, за время π точка проходит половину окружности и занимает положение, диаметрально противоположное исходному.

Свойство 4

Для всякого значения t точки P_t и P_{-t} симметричны друг другу относительно оси абсцисс (рис. 139).

Действительно, для построения точек P_t и P_{-t} надо отложить от точки P_0 дуги, равные $|t|$, но в противоположных направлениях.

Свойство 5

Для всякого значения t точки P_t и $P_{-t+\pi}$ симметричны относительно оси ординат (рис. 140—142).

Это свойство выводится из свойств 3, 4.

Свойство 6

Для всякого значения t точки P_t и $P_{\pi/2-t}$ симметричны друг другу относительно биссектрисы первого и третьего координатных углов (рис. 143).

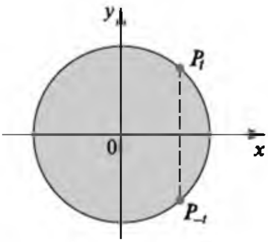


Рис. 140

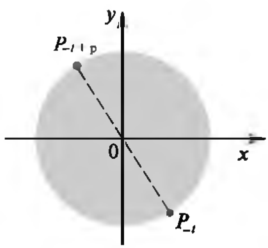


Рис. 141

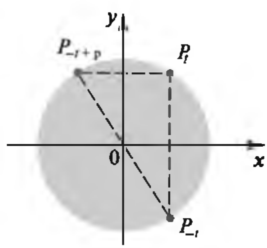


Рис. 142

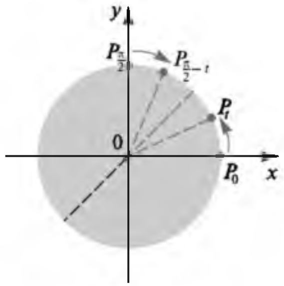


Рис. 143

Возьмем две точки: P_0 и $P_{\pi/2}$. Они симметричны друг другу относительно биссектрисы первого и третьего координатных углов. Точка P_t получится при перемещении на расстояние $|t|$ в одном каком-то направлении от P_0, а точка $P_{\pi/2-t}$ — при перемещении на такое же расстояние от точки $P_{\pi/2}$, но в противоположном направлении.

В этом случае точки P_t и $P_{\pi/2-t}$ при всяком t будут оставаться симметричными друг другу относительно указанной прямой.

Контрольные вопросы и задания

1. Обратите внимание на следующие ключевые слова и обозначения, появившиеся в этом параграфе: угол, радиан, число π, точка P_t. Приведите примеры их использования.
2. По каким формулам переводят градусную меру угла в радианную и наоборот?
3. Как построить произвольный угол t?
4. В чем состоит периодичность вращательного движения?
5. Каким углам соответствуют диаметрально противоположные точки окружности?
6. Каким углам соответствуют точки, симметричные друг другу относительно оси абсцисс?
7. Каким углам соответствуют точки, симметричные относительно оси ординат?
8. Каким углам соответствуют точки, симметричные друг другу относительно прямой $y = x$?
9. Как связаны между собой углы t и t', если $P_t = P_{t'}$?
10. Приведите примеры периодических процессов.

§ 13. Определение тригонометрических функций

Тригонометрия треугольника

Тригонометрия возникла как наука об измерении треугольников в связи с необходимостью производить астрономические вычисления и долгое время была частью астрономии. В средние века она отделилась от астрономии и стала самостоятельной частью математики. Последовательное изучение тригонометрических функций как функций произвольного числового аргумента связано с именем замечательного математика XVIII в. Леонарда Эйлера. Оно открыло дорогу самым широким применениям тригонометрических функций и вместе с тем соединило их изучение с изучением других функций, покончив с изолированным положением тригонометрии внутри математики.

Вычисления в треугольниках — это истоки тригонометрических функций. Значение этих вычислений сохраняется до наших дней. Поэтому мы начнем изучение тригонометрии с того, что вспомним, как применяются синус, косинус, тангенс и котангенс для решения треугольников.

В треугольнике три угла. Их можно измерять в градусах. Сумма углов треугольника равна 180° (рис. 144). Угол 90° называется *прямым*, меньший 90° — *острым*, больший 90° — *тупым* (рис. 145).

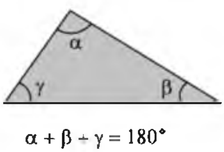


Рис. 144

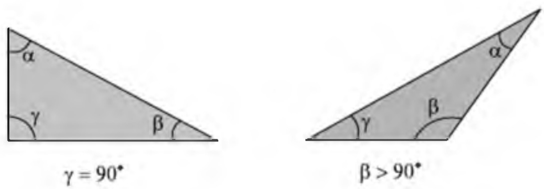


Рис. 145

В прямоугольном треугольнике отношения сторон не зависят от их длин, а зависят только от углов треугольника (рис. 146). Эти отношения называются синусом, косинусом, тангенсом и котангенсом.

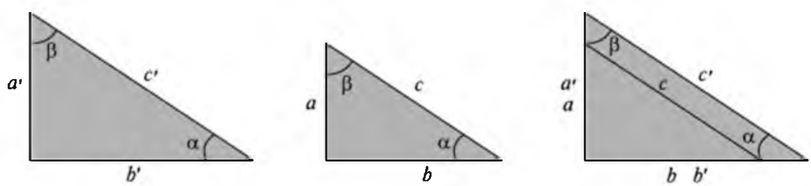


Рис. 146

Пусть α — острый угол прямоугольного треугольника. Отношение противолежащего ему катета к гипотенузе равно синусу угла α, отношение прилежащего катета к гипотенузе равно косинусу угла α. Отношение противолежащего катета к прилежащему равно тангенсу, прилежащего катета к противолежащему — котангенсу.

Основные соотношения для прямоугольного треугольника (рис. 147):

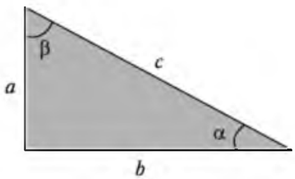


Рис. 147

$$\sin\alpha = \frac{a}{c};\ \cos\alpha = \frac{b}{c};$$

$$\operatorname{tg}\alpha = \frac{a}{b};\ \operatorname{ctg}\alpha = \frac{b}{a};$$

$\sin^2\alpha + \cos^2\alpha = 1$ (теорема Пифагора);

$$\operatorname{tg}\alpha = \frac{\sin\alpha}{\cos\alpha};\ \operatorname{ctg}\alpha = \frac{\cos\alpha}{\sin\alpha};$$

$$\operatorname{tg}\alpha\operatorname{ctg}\alpha = 1;$$

$$\sin\beta = \cos\alpha;\ \cos\beta = \sin\alpha;\ \operatorname{tg}\beta = \operatorname{ctg}\alpha;\ \operatorname{ctg}\beta = \operatorname{tg}\alpha.$$

Примеры

1. Равнобедренный прямоугольный треугольник. Его острые углы $\alpha = \beta = 45°$ (рис. 148). Оба катета в таком треугольнике равны, и если их длины принять равными 1, то по теореме Пифагора, длина гипотенузы равна $\sqrt{2}$. Поэтому

$$\sin 45° = \cos 45° = \frac{\sqrt{2}}{2};\ \operatorname{tg} 45° = \operatorname{ctg} 45° = 1.$$

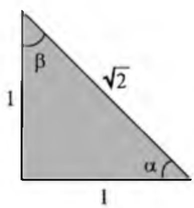


Рис. 148

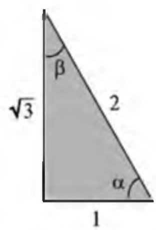


Рис. 149

2. Прямоугольный треугольник, у которого гипотенуза вдвое больше одного из катетов. Его острые углы — 30° и 60° (рис. 149). Катет, лежащий против угла 30° в таком треугольнике, в два раза меньше гипотенузы. Если его длину принять равной 1 (т.е. гипотенуза равна 2), то другой катет, по теореме Пифагора, равен $\sqrt{3}$. Поэтому

$$\sin 30° = \cos 60° = \frac{1}{2};\ \operatorname{tg} 30° = \operatorname{ctg} 60° = \frac{1}{\sqrt{3}} = \frac{\sqrt{3}}{3};$$

$$\sin 60° = \cos 30° = \frac{\sqrt{3}}{2};\ \operatorname{ctg} 30° = \operatorname{tg} 60° = \sqrt{3}.$$

В общем случае знание одной из сторон и одной из тригонометрических функций острого угла достаточно для вычисления всех элементов прямоугольного треугольника. Знание двух сторон (а значит, и третьей, по теореме Пифагора) достаточно для вычисления тригонометрических функций углов прямоугольного треугольника.

Координаты вращающейся точки

Первоначально тригонометрические функции — синус, косинус, тангенс и котангенс — были введены нами для острых углов — углов прямоугольного треугольника. Определим эти функции для произвольного числового значения аргумента, который пока обозначим буквой t, так как тригонометрические функции связаны с некоторым движением, зависящим от времени t.

Рассмотрим на координатной плоскости xOy единичную окружность, т.е. окружность единичного радиуса с центром в начале координат. Обозначим через A точку единичной окружности с координатами (1, 0) (рис. 150). Точку A будем называть ***начальной точкой***. Возьмем произвольное число t. Повернем начальную точку на угол t. Получим точку на единичной окружности, которую обозначим P_t.

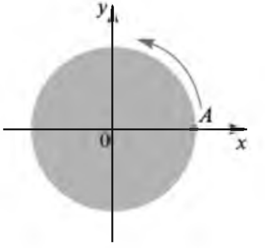


Рис. 150

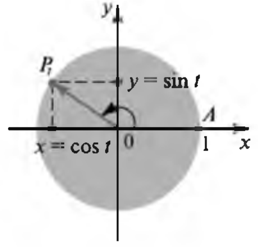


Рис. 151

Синусом числа t называется ордината точки P_t: $\sin t = y$.

Косинусом числа t называется абсцисса точки P_t: $\cos t = x$ (рис. 151).

Тангенсом числа t называется отношение ординаты точки P_t к ее абсциссе: $\operatorname{tg} t = \dfrac{y}{x}$.

Котангенсом числа t называется отношение абсциссы точки P_t к ее ординате: $\operatorname{ctg} t = \dfrac{x}{y}$.

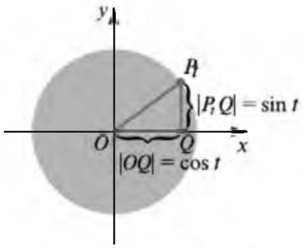


Рис. 152

Если угол t лежит в первой четверти, то абсцисса и ордината точки P_t являются длинами катетов треугольника OP_tQ (рис. 152). В этом случае определения тригонометрических функций являются теми же самыми, какими они были для углов прямоугольного треугольника. В остальных четвертях значения тригонометрических функций также могут быть определены из треугольников, только при этом надо учитывать знаки координат точки P_t и связь угла t с острыми углами треугольника.

Периодичность

При изучении вращательного движения в качестве первого его свойства мы назвали периодичность; пройдя полный оборот, вращающаяся точка займет то же положение, что и вначале:

$$P_{t+2\pi k} = P_t \ (k \in \mathbb{Z}).$$

Тригонометрические функции определены с помощью координат вращающейся точки. Поэтому если для двух чисел t и $t + 2\pi k$ положения вращающейся точки совпадают, то совпадают и значения всех тригонометрических функций от этих чисел:

$$\sin (t + 2\pi k) = \sin t; \cos (t + 2\pi k) = \cos t; \operatorname{tg} (t + 2\pi k) = \operatorname{tg} t;$$
$$\operatorname{ctg} (t + 2\pi k) = \operatorname{ctg} t.$$

Функции, значения которых не изменяются при прибавлении к аргументу какого-либо постоянного числа, называются ***периодическими функциями***. Синус, косинус, тангенс и котангенс являются периодическими функциями.

Равенство $\sin (t + 2\pi) = \sin t$ верно при всех значениях t. Это означает, что число 2π является периодом синуса. Подставляя в это равенство вместо t число $t + 2\pi$, получаем цепочку равенств $\sin (t + 2\pi + 2\pi) = \sin (t + 2\pi) = \sin t$, т.е. равенство $\sin (t + 4\pi) = \sin t$ также верно при всех значениях t. Аналогично, подставляя вместо t число $t - 2\pi$, получаем тождество $\sin (t - 2\pi) = \sin t$. Так как 2π — период синуса, то и $2 \cdot 2\pi$, -2π, ... также являются его периодами. Точно так же всякое число вида $2\pi k$ ($k \in \mathbb{Z}$) являются периодом синуса.

Однако 2π выделяется тем, что это наименьший положительный период синуса. Аналогично, 2π — наименьший положительный период косинуса. В отличие от них 2π хотя и является периодом тангенса и котангенса, однако не наименьшим среди положительных периодов этих функций.

Знаки тригонометрических функций

Знаки тригонометрических функций определяются в зависимости от того, в какой четверти лежит рассматриваемый угол.

Синус числа t есть ордината точки P_t. Поэтому синус положителен в первой и второй четвертях и отрицателен в третьей и четвертой (рис. 153).

Аналогично, косинус числа t как абсцисса точки P_t положителен в первой и четвертой четвертях и отрицателен во второй и третьей (рис. 154).

Знаки синуса

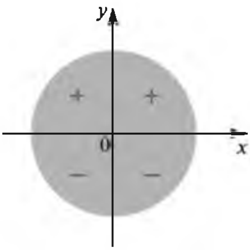


Рис. 153

Знаки косинуса

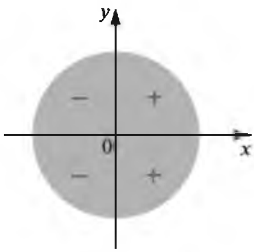


Рис. 154

Четность

Синус — нечетная функция: при всех t выполнено равенство $\sin(-t) = -\sin t$ (рис. 155).

Косинус — четная функция: при всех t выполнено равенство $\cos(-t) = \cos t$.

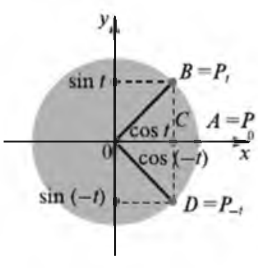


Рис. 155

Действительно, для всякого значения t точки P_t и P_{-t} симметричны друг другу относительно оси абсцисс (свойство 4 вращательного движения). У точек P_t и P_{-t} совпадают абсциссы ($\cos(-t) = \cos t$), а ординаты противоположны ($\sin(-t) = -\sin t$).

Тангенс и котангенс — нечетные функции

Действительно, известно, как эти функции выражаются через синус и косинус:

$$\operatorname{tg} t = \frac{y}{x} = \frac{\sin t}{\cos t}, \ \operatorname{ctg} t = \frac{x}{y} = \frac{\cos t}{\sin t},$$

Зная поведение синуса и косинуса при замене t на $-t$, получаем доказательство нечетности тангенса (и аналогично, котангенса):

$$\operatorname{tg}(-t) = \frac{\sin(-t)}{\cos(-t)} = \frac{-\sin t}{\cos t} = -\operatorname{tg} t.$$

Простейшие тождества

Тригонометрические функции связаны между собой рядом тождеств.

Основные тождества:

1. $\sin^2 t + \cos^2 t = 1$;
2. $\operatorname{tg} t = \dfrac{\sin t}{\cos t}$;
3. $\operatorname{ctg} t = \dfrac{\cos t}{\sin t}$;
4. $\operatorname{tg} t \operatorname{ctg} t = 1$;
5. $1 + \operatorname{tg}^2 t = \dfrac{1}{\cos^2 t}$;
6. $1 + \operatorname{ctg}^2 t = \dfrac{1}{\sin^2 t}$.

Тождество 1 является следствием теоремы Пифагора. Тождества 2 и 3 — это другая запись определения тангенса и котангенса, которую мы уже использовали раньше. Тождество 4 получается перемножением тождеств 2 и 3. Покажем вывод тождества 5:

$$1 + \operatorname{tg}^2 t = 1 + \frac{\sin^2 t}{\cos^2 t} = \frac{\cos^2 t + \sin^2 t}{\cos^2 t} = \frac{1}{\cos^2 t}.$$

Формулы приведения

Значения тригонометрических функций острых углов можно найти по таблицам или с помощью прямоугольного треугольника. Их вычисление для любого значения аргумента можно привести к вычислению значений для аргумента $t \in \left[0, \frac{\pi}{2}\right]$. Соответствующие формулы и называются — *формулы приведения*. Они основаны на симметрии вращательного движения.

Основные формулы:

1. $\sin(t+\pi) = -\sin t$, $\cos(t+\pi) = -\cos t$.
2. $\sin(\pi - t) = \sin t$, $\cos(\pi - t) = -\cos t$.
3. $\sin\left(\frac{\pi}{2} - t\right) = \cos t$, $\cos\left(\frac{\pi}{2} - t\right) = \sin t$.

Формулы 1 — это запись в координатной форме свойства 3 вращательного движения, формулы 2 — свойства 5, а формулы 3 — свойства 6.

Используя свойства периодичности, четности и формулы 1—3, можно привести вычисление синуса и косинуса любого числа к значению t, лежащему между 0 и $\frac{\pi}{2}$, т.е. к $0 < t \le \frac{\pi}{2}$. Поэтому формулы 1—3 называют формулами приведения.

Можно вывести и другие формулы приведения:

$$\sin\left(\frac{\pi}{2} + t\right) = \cos t; \ \cos\left(\frac{\pi}{2} + t\right) = -\sin t.$$

Доказательство:

$$\sin\left(\frac{\pi}{2} + t\right) = \sin\left(\frac{\pi}{2} - (-t)\right) = \cos(-t) = \cos t;$$

$$\cos\left(\frac{\pi}{2} + t\right) = \cos\left(\frac{\pi}{2} - (-t)\right) = \sin(-t) = -\sin t.$$

Аналогично выведем формулы $\sin\left(\frac{3\pi}{2} + t\right) = -\cos t$; $\cos\left(\frac{3\pi}{2} + t\right) = -\sin t$.

Формулы приведения для тангенса и котангенса получаются как следствие полученных формул для синуса и косинуса, например

$$\operatorname{ctg}\left(\frac{\pi}{2} + t\right) = \frac{\cos\left(\frac{\pi}{2} + t\right)}{\sin\left(\frac{\pi}{2} + t\right)} = \frac{-\sin t}{\cos t} = -\operatorname{tg} t.$$

Для применения формул приведения полезно запомнить следующее правило.

Название функции не меняется, если к аргументу левой части добавляется $-\pi$ или $+\pi$, и меняется, если добавляются числа $\pm\frac{\pi}{2}$ или $\pm\frac{3\pi}{2}$.

Знак в правой части определяется знаком левой при $0 < t < \frac{\pi}{2}$.

Примеры

1. Вычислить $\sin\frac{41\pi}{6}$.

   $\frac{41\pi}{6} = 6\pi + \frac{5}{6}\pi$. Далее, $\sin\frac{41\pi}{6} = \sin\left(6\pi+\frac{5}{6}\pi\right) = \sin\frac{5}{6}\pi =$ $= \sin\left(\pi-\frac{\pi}{6}\right) = \sin\frac{\pi}{6} = \frac{1}{2}$.

2. $\cos\frac{82\pi}{3} = \cos\left(26\pi+\frac{4\pi}{3}\right) = \cos\frac{4\pi}{3} = \cos\left(\pi+\frac{\pi}{3}\right) = -\cos\frac{\pi}{3} = -\frac{1}{2}$.

   А можно иначе: $\cos\frac{82\pi}{3} = \cos\left(28\pi-\frac{2\pi}{3}\right) = \cos\left(-\frac{2\pi}{3}\right) = \cos\frac{2\pi}{3} =$ $= \cos\left(\pi-\frac{\pi}{3}\right) = -\frac{1}{2}$.

Значения тригонометрических функций

Вычисление значений тригонометрических функций имеет длинную историю. Потребности точных астрономических наблюдений вызвали к жизни появление огромных таблиц, позволяющих производить вычисления с четырьмя, пятью и даже семью (и более) знаками. На составление этих таблиц было потрачено много усилий. Сейчас, нажав на кнопку микрокалькулятора, можно моментально получить требуемое значение величины с очень высокой точностью. С помощью ЭВМ нетрудно найти значения тригонометрических функций с еще большей степенью точности.

Известные нам свойства тригонометрических функций облегчают вычисления и позволяют разумно использовать вычислительные устройства.

1. С помощью формул приведения вычисление значения тригонометрической функции любого числа можно свести к вычислению функции от угла, лежащего в первой четверти.
2. Достаточно знать значение лишь одной из тригонометрических функций. С помощью основных тождеств и зная четверть, в которой лежат значения аргумента, легко найти значения остальных функций.

Примеры

1. Пусть $\sin t = -\frac{3}{5}$ и t лежит в третьей четверти.
Тогда $\cos^2 t = 1 - \sin^2 t = \frac{16}{25}$ и $\cos t = -\frac{4}{5}$, так как косинус в третьей четверти отрицателен. Получаем $\operatorname{tg} t = \frac{3}{4}$ и $\operatorname{ctg} t = \frac{4}{3}$.
2. $\cos t = \frac{1}{3}$, $t \in \left[\frac{3\pi}{2}; 2\pi\right]$. Получаем: $\sin t = -\sqrt{1-\frac{1}{9}} = -\frac{2\sqrt{2}}{3}$;
$\operatorname{tg} t = -2\sqrt{2}$; $\operatorname{ctg} t = -\frac{1}{2\sqrt{2}} = -\frac{\sqrt{2}}{4}$.
3. $\operatorname{tg} t = -10$ и t лежит во второй четверти. Получаем: $\cos^2 t = \frac{1}{1+\operatorname{tg}^2 t} = \frac{1}{101}$ и $\cos t = -\frac{1}{\sqrt{101}}$; $\sin t = \operatorname{tg} t \cos t = \frac{10}{\sqrt{101}}$;
$\operatorname{ctg} t = -\frac{1}{10}$.

Таблица 2

Значения тригонометрических функций

t	$\sin t$	$\cos t$	$\operatorname{tg} t$	$\operatorname{ctg} t$
0	0	1	0	—
$\frac{\pi}{6}$	$\frac{1}{2}$	$\frac{\sqrt{3}}{2}$	$\frac{1}{\sqrt{3}} = \frac{\sqrt{3}}{3}$	$\sqrt{3}$
$\frac{\pi}{4}$	$\frac{1}{\sqrt{2}} = \frac{\sqrt{2}}{2}$	$\frac{1}{\sqrt{2}} = \frac{\sqrt{2}}{2}$	1	1
$\frac{\pi}{3}$	$\frac{\sqrt{3}}{2}$	$\frac{1}{2}$	$\sqrt{3}$	$\frac{1}{\sqrt{3}} = \frac{\sqrt{3}}{3}$
$\frac{\pi}{2}$	1	0	—	0

Полезно помнить значения тригонометрических функций для углов двух прямоугольных треугольников: равнобедренного и треугольника с углами 30° $\left(\frac{\pi}{6}\right)$ и 60° $\left(\frac{\pi}{3}\right)$. Эти значения обычно записывают с помощью радикалов и при необходимости эти радикалы заменяют их приближенными значениями ($\sqrt{2} = 1{,}414...$, $\sqrt{3} = 1{,}7...$). Сведем их в табл. 2, дополнив ее значениями тригонометрических функций аргументов $t = 0$ и $t = \frac{\pi}{2}$.

Решение простейших уравнений

Мы рассмотрели вопрос о вычислении значений тригонометрических функций. Решение уравнений приводит к обратной задаче — по значению функции найти угол. Для полного решения этой задачи необходимо ввести обратные тригонометрические функции, изучение которых мы отложим на будущее.

Для решения некоторых особенно простых, но важных уравнений достаточно вспомнить определение тригонометрических функций с помощью координат точек единичной окружности.

Примеры

1. $\sin t = 0$. Вращающаяся точка P_t имеет нулевую ординату в моменты времени $t = 0, \pi, 2\pi, \ldots$, а также $t = -\pi, -2\pi, \ldots$ В общем виде множество этих значений можно записать в виде $t = \pi k$ $(k \in \mathbb{Z})$. Таким образом, решением уравнения $\sin t = 0$ будут числа $t = \pi k$ (рис. 156).

   Запишем кратко ответы для решения еще нескольких уравнений, которые вам предлагаются для самостоятельной проверки $(k \in \mathbb{Z})$.

2. $\sin t = 1$, $t = \dfrac{\pi}{2} + 2\pi k$ (рис. 157).
3. $\sin t = -1$, $t = -\dfrac{\pi}{2} + 2\pi k$ (рис. 158).
4. $\cos t = 0$, $t = \dfrac{\pi}{2} + \pi k$ (рис. 159).
5. $\cos t = 1$, $t = 2\pi k$ (рис. 160).
6. $\cos t = -1$, $t = \pi + 2\pi k$ (рис. 161).

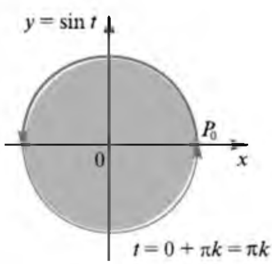


Рис. 156

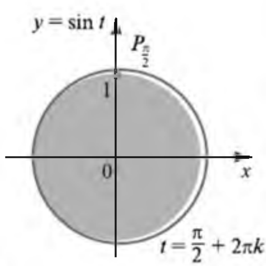


Рис. 157

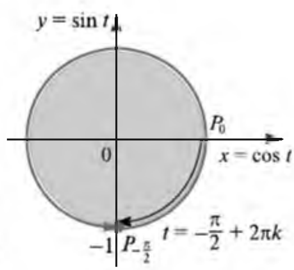


Рис. 158

Заметим, что все рассмотренные уравнения имеют бесчисленное множество решений. Они записываются в виде бесконечных серий с помощью переменной (в наших примерах — k), которая может принимать любые целые значения.

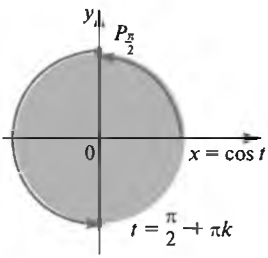


Рис. 159

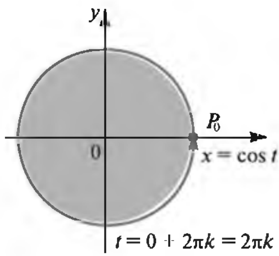


Рис. 160

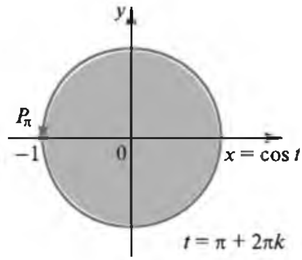


Рис. 161

Часто приходится искать не все решения тригонометрического уравнения, а только те, которые лежат в определенном промежутке. Тогда решений обычно бывает конечное число (часто просто одно) и их можно перечислить.

Примеры

1. $\cos t = 0,\ t \in [\pi, 2\pi]$.
   В указанном промежутке уравнение $\cos t = 0$ имеет одно решение $t = \dfrac{3\pi}{2}$.
2. $\sin t = \dfrac{1}{2},\ t \in \left[\dfrac{\pi}{2}, \pi\right]$.

Из рисунка 162 видно, что решение единственно: $t = \pi - \dfrac{\pi}{6} = \dfrac{5\pi}{6}$.

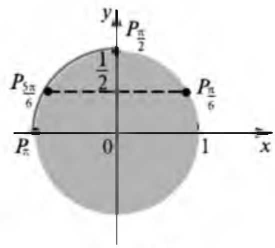


Рис. 162

Контрольные вопросы и задания

1. Обратите внимание на следующие ключевые слова и выражения, появившиеся в этом параграфе: синус, косинус, тангенс, котангенс, периодичность, формулы приведения. Приведите примеры их употребления.
2. Дайте определение основных тригонометрических функций.
3. Какое число является наименьшим положительным периодом синуса и косинуса?
4. Определите знаки тригонометрических функций в зависимости от того, в какой четверти находится угол.
5. Что можно сказать о четности тригонометрических функций?
6. Какое вы знаете правило для запоминания формул приведения?

7. Назовите значения тригонометрических функций для углов 30°, 45°, 60°.
8. Какова область определения тангенса?
9. При каких углах котангенс угла не определен?
10. Значение $\sin t$ вам известно. Достаточно ли этого, чтобы найти значения других тригонометрических функций угла t?

§ 14. Исследование тригонометрических функций

Основные свойства синуса и косинуса

При введении тригонометрических функций мы аргумент обозначили буквой t, так как буквы x и y обозначали координаты вращающейся точки P_t. Сейчас при исследовании тригонометрических функций мы вернемся к обычным обозначениям: x — аргумент, y — функция.

Рассмотрим свойства функции $y = \sin x$ и $y = \cos x$.

1. Область определения. Синус и косинус числа x задаются как координаты точки P_x, получающейся из точки (1, 0) поворотом на угол x. Так как поворот возможен на любой угол, то

областью определения синуса и косинуса является множество всех вещественных чисел, т.е. $D = \mathbb{R}$.

2. Периодичность. Мы знаем, что 2π является периодом синуса и косинуса. Докажем, что

2π — наименьший положительный период синуса и косинуса.

Докажем это утверждение для синуса. Пусть число $T > 0$ такое, что равенство $\sin(x + T) = \sin x$ выполняется тождественно. При $x = 0$ получим $\sin T = \sin 0 = 0$. Есть лишь одно положительное число T, меньшее 2π, для которого $\sin T = 0$ — это число $T = \pi$. Следовательно, единственно возможным положительным периодом синуса, меньшим 2π, является число π. Однако оно не является периодом, так как $\sin(x + \pi) = -\sin x$ (по формуле приведения), а $(-\sin x)$ не может быть равным тождественно $\sin x$. Утверждение доказано.

3. Участки монотонности. Проследим за характером изменения координат точки, движущейся по окружности. При $x = 0$ точка занимает положение $P_0(1, 0)$ (рис. 163). Пока она движется по окружности, оставаясь в первой четверти, ее абсцисса уменьшается, а ордината увеличивается. При $x = \frac{\pi}{2}$ точка займет положение $P_{\frac{\pi}{2}}(0, 1)$. Итак,

в первой четверти синус (ордината) возрастает от 0 до 1, а косинус (абсцисса) убывает от 1 до 0.

Когда точка переходит во вторую четверть (рис. 164), ордината начинает убывать и меняется от 1 до 0. Абсцисса становится отрицательной и растет по абсолютной величине — значит, косинус продолжает убывать от 0 до -1.

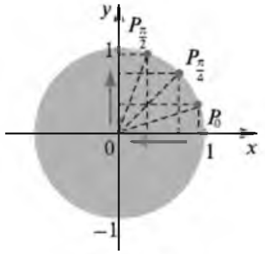


Рис. 163

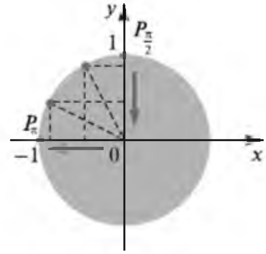


Рис. 164

В третьей четверти синус становится отрицательным и убывает от 0 до -1 (рис. 165), а косинус начинает возрастать и изменяется от -1 до 0.

Наконец, в четвертой четверти синус возрастает от -1 до 0 и косинус возрастает от 0 до 1 (рис. 166).

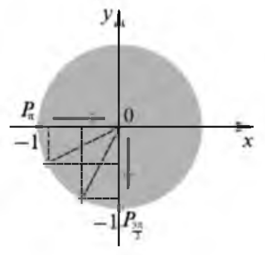


Рис. 165

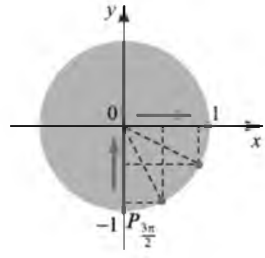


Рис. 166

Составим таблицу 3 для определения монотонности синуса и косинуса по четвертям.

Таблица 3

Монотонность синуса и косинуса

Функция	Четверть			
	I	II	III	IV
sin	↗	↘	↘	↗
cos	↘	↘	↗	↗

4. Точки экстремума. Координаты вращающейся точки меняются между -1 и $+1$. Эти числа являются ***наименьшим и наибольшим значениями синуса и косинуса***. Если требуется указать абсциссы точек экстремума, то надо решить уравнения $\sin x = \pm 1$ и $\cos x = \pm 1$ (см. § 13).
5. Участки знакопостоянства и точки обращения в нуль. Эти вопросы были нами также изучены в § 13. Мы повторим их еще раз при построении графика.
6. Множество значений. Синус и косинус принимают любые значения от -1 до $+1$, так как являются координатами точки, движущейся по единичной окружности.

Графики синуса и косинуса

Для приближенного построения графика синуса — синусоиды можно поступить так. Разделим первую четверть на восемь равных частей и на столько же частей разделим отрезок $\left[0; \frac{\pi}{2}\right] = [0;\ 1{,}57...]$ оси абсцисс. Удобно при этом сместить окружность влево (рис. 167). Перенесем проекции на ось y точек деления окружности (значения синуса) к соответствующим точкам деления отрезка $\left(0; \frac{\pi}{2}\right)$. Получим точки, лежащие на синусоиде, которые нужно плавно соединить и продолжить кривую дальше, пользуясь симметрией.

Будем строить график функции $y = \sin x$, учитывая свойства синуса.

Мы получили график синуса на промежутке $\left[0; \frac{\pi}{2}\right]$. Так как $\sin\left(\frac{\pi}{2} - x\right) = \sin\left(\frac{\pi}{2} + x\right)$, то график синуса симметричен относительно прямой $x = \frac{\pi}{2}$. Это позволяет построить график синуса на отрезке $\left[\frac{\pi}{2}; \pi\right]$.

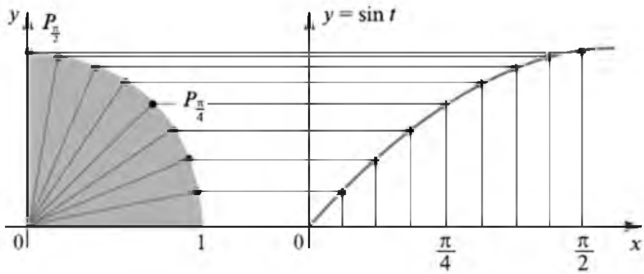


Рис. 167

Функция $y = \sin x$ нечетна.

На графике это свойство проявляется так — синусоида симметрична относительно начала координат (рис. 168).

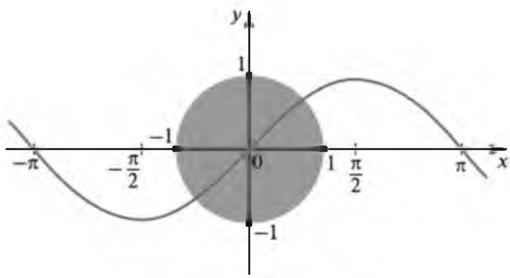


Рис. 168

Воспользовавшись нечетностью синуса, получим график синуса на отрезке $[-\pi; 0]$ симметричным отражением построенной части синусоиды относительно начала координат.

Функция $y = \sin x$ имеет период 2π.

Так как отрезок $[-\pi; \pi]$ имеет длину, равную периоду синуса, то график синуса на всей числовой оси можно получить параллельными переносами построенной кривой.

На графике это свойство отражается следующим образом: если мы разобьем ось x на отрезки длиной 2π, например точками -4π, -2π, 0, 2π, 4π, ..., то весь график разобьется на «одинаковые» части, получающиеся друг из друга параллельным переносом вдоль оси (рис. 169).

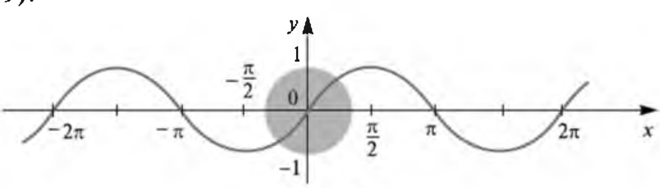


Рис. 169

График синуса мы построили, воспользовавшись его свойствами. При этом к определению синуса мы обращались только при построении графика на отрезке $\left[0; \frac{\pi}{2}\right]$.

Построение графика на всей оси потребовало знания симметрии вращательного движения (формулы приведения, нечетность, периодичность). После того как график построен, полезно вернуться к свойствам синуса и посмотреть, как они проявляются на графике.

Функция $y = \sin x$ обращается в нуль при $x = k\pi$ ($k \in \mathbb{Z}$).

На графике — это точки пересечения синусоиды с осью абсцисс.

Функция $y = \sin x$ положительна при $2k\pi < x < (2k + 1)\pi$ и отрицательна при $(2k + 1)\pi < x < (2k + 2)\pi$ ($k \in \mathbb{Z}$).

Указанные отрезки соответствуют тем значениям угла x, которые лежат в первой-второй (где $\sin x > 0$) или третьей-четвертой четвертях (где $\sin x < 0$) (рис. 170).

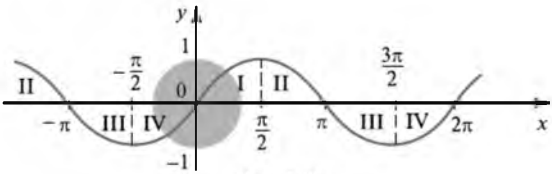


Рис. 170

Функция $y = \sin x$ возрастает при $-\frac{\pi}{2} + 2\pi k \le x \le \frac{\pi}{2} + 2\pi k$ и убывает при $\frac{\pi}{2} + 2\pi k \le x \le \frac{3\pi}{2} + 2\pi k$. Точно так же указанные отрезки соответствуют четвертой-первой и второй-третьей четвертям.

Множеством значений функции $y = \sin x$ является отрезок $[-1; 1]$.

Действительно, проекции движущейся точки P на ось y заполняют отрезок $[-1; 1]$ (рис. 171). Следовательно, синусоида расположена в полосе $-1 \le \sin x \le 1$ и при этом проекции точек графика целиком заполняют отрезок $[-1; 1]$ (рис. 172).

График косинуса можно построить так же, как и график синуса (рис. 173). Но возможен и другой путь. Формулы приведения показывают, что синус и косинус связаны между собой простыми соотношениями. Воспользуемся, например, формулой $\cos x = \sin\left(x + \frac{\pi}{2}\right)$. Эта формула показывает, что график косинуса получается сдвигом синусоиды на $\frac{\pi}{2}$ влево по оси x (рис. 174).

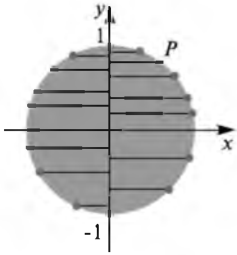


Рис. 171

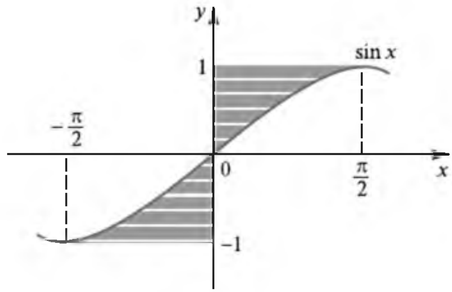


Рис. 172

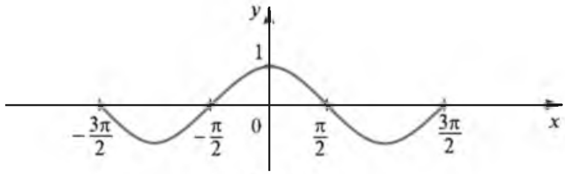


Рис. 173

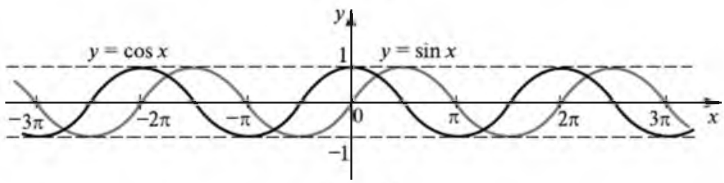


Рис. 174

Если аргумент умножать на число, то синусоида будет сжиматься или стягиваться по оси x (рис. 175). Если значения синуса умножить на число, то произойдет растяжение (сжатие) синусоиды по оси y (рис. 176).

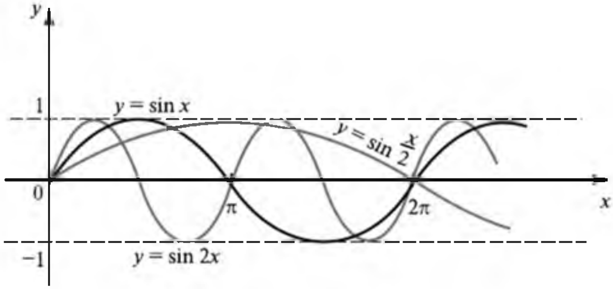


Рис. 175

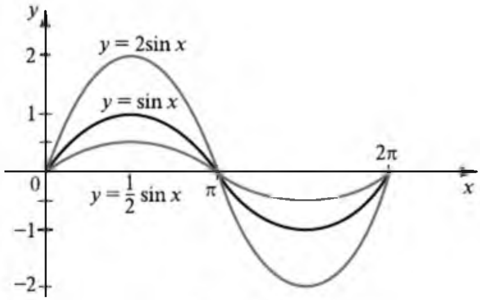


Рис. 176

Гармонические колебания

Гармоническое колебание — это процесс, который может быть описан функцией вида

$$y = A\sin(\omega x + \alpha).$$

Примеры

1. *Колебания упругой пружины.* Конец упругой пружины (точка P; рис. 177) при ее сжатии или растяжении описывает колебательные движения. Если на прямой, по которой движется точка P, ввести координату x так, что в положении равновесия $x_P = 0$, оттянуть конец пружины в положительном направлении на расстояние A и в момент времени $t = 0$ отпустить его, то зависимость координаты точки P от времени t имеет вид $x = A\cos\omega t = A\sin\left(\omega t + \frac{\pi}{2}\right)$, где ω — некоторый коэффициент, характеризующий упругость пружины.

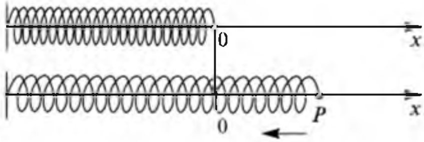


Рис. 177

2. *Электрический колебательный контур.* Рассмотрим электрическую цепь, состоящую из последовательно соединенных конденсатора и катушки индуктивности (рис. 178). Если эту цепь замкнуть накоротко и считать, что в ней есть некоторый запас

энергии (например, ненулевой заряд в конденсаторе), то по этой цепи пойдет ток, напряжение U которого меняется со временем. При идеальном предположении отсутствия потерь в цепи зависимость U от времени t имеет вид

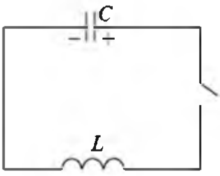


Рис. 178

$$U = U_0 \sin(\omega t + \alpha),$$

где U_0 — некоторая характеристика контура, которая вычисляется через параметры конденсатора и катушки. Константы U_0 и α зависят от состояния цепи в начальный момент времени.

Таким образом, гармоническое колебание $y = A\sin(\omega t + \alpha)$ определяется тремя параметрами: амплитудой $A > 0$, угловой скоростью $\omega > 0$ и начальной фазой α. Часто вместо угловой скорости ω говорят о частоте колебаний ν, которая связана с угловой скоростью ω (круговой частотой) формулой

$$\omega = 2\pi\nu.$$

Функция y периодична. Ее основной период равен

$$T = \frac{2\pi}{\omega} = \frac{1}{\nu}.$$

Исследование тангенса

Если свойства синуса и косинуса мы получили, рассматривая свойства движения точки по окружности, то для исследования тангенса и котангенса поступим иначе.

По определению, тангенс числа x задается как отношение $\sin x$ к $\cos x$. Рассмотрим свойства тангенса.

1. Область определения

Областью определения функции

$$y = \operatorname{tg} x = \frac{\sin x}{\cos x}$$

является множество всех действительных чисел, за исключением тех, в которых косинус обращается в нуль:

$$\left\{x \neq \frac{\pi}{2} + \pi k, (k \in Z)\right\}.$$

2. Периодичность

Тангенс — периодическая функция с периодом π.
π — наименьший положительный период тангенса и котангенса.

Проверим, что π является периодом этих функций:

$$\operatorname{tg}(x+\pi)=\frac{\sin(x+\pi)}{\cos(x+\pi)}=\frac{-\sin x}{-\cos x}=\operatorname{tg} x.$$

Для котангенса доказательство проводится аналогично.

То, что π — наименьший период, доказывается легко: из тождества $\operatorname{tg}(x+T)=\operatorname{tg} x$ при $x=0$ следует, что $\operatorname{tg} T=0$, а π — наименьший положительный корень этого уровнения. Так как $\operatorname{ctg} x=\dfrac{1}{\operatorname{tg} x}$, то аналогичное утверждение верно и для котангенса.

3. Нечетность

Тангенс — нечетная функция: $\operatorname{tg}(-x)=-\operatorname{tg}(x)$

4. Нахождение нулей

Функция $y=\operatorname{tg} x$ обращается в нуль одновременно с синусом, т.е. при $x=k\pi$ ($k\in\mathbb{Z}$).

5. Знаки

Функция $y=\operatorname{tg} x$ положительна в первой и третьей четвертях и отрицательна во второй и четвертой (рис. 179)

Выберем для дальнейшего изучения тангенса какой-либо отрезок числовой оси с длиной, равной периоду, т.е. числу π. Можно было бы использовать отрезок от 0 до π, но это неудобно, так как внутри этого отрезка есть точка $x=\dfrac{\pi}{2}$, в которой тангенс не определен. Лучше выбрать отрезок $\left(-\dfrac{\pi}{2},\dfrac{\pi}{2}\right)$.

Знаки тангенса и котангенса

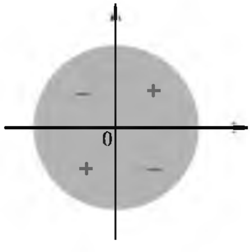


Рис. 179

6. Монотонность

Тангенс возрастает в первой четверти.

Действительно, пусть $0\le x_1<x_2<\dfrac{\pi}{2}$. Тогда $\sin x_1<\sin x_2$ (возрастание синуса) и $\cos x_1>\cos x_2$ (убывание косинуса). Так как значения косинуса положительны, то, по свойству неравенств,

$$\frac{1}{\cos x_1} < \frac{1}{\cos x_2}.$$

Умножив это неравенство на неравенство $\sin x_1 < \sin x_2$ с положительными членами, получим $\operatorname{tg} x_1 < \operatorname{tg} x_2$. *Тангенс возрастает также и в четвертой четверти*. Действительно, пусть $-\frac{\pi}{2} < x_1 < x_2 \le 0$. Тогда $0 \le -x_2 < -x_1 < \frac{\pi}{2}$. Теперь числа $(-x_1)$ и $(-x_2)$ лежат в первой четверти и можно воспользоваться тем, что в первой четверти тангенс возрастает: $\operatorname{tg}(-x_2) < \operatorname{tg}(-x_1)$. Но так как тангенс — нечетная функция, то

$$\operatorname{tg}(-x_2) < \operatorname{tg}(-x_1) \Leftrightarrow -\operatorname{tg} x_2 < -\operatorname{tg} x_1 \Leftrightarrow \operatorname{tg} x_1 < \operatorname{tg} x_2.$$

От $-\frac{\pi}{2}$ до 0 тангенс отрицателен и возрастает. Затем от 0 до $\frac{\pi}{2}$ тангенс положителен и возрастает. Итак,

тангенс возрастает на отрезке $\left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$.

7. Область значений

Какие же значения принимает тангенс? Когда x возрастает от 0 до $\frac{\pi}{2}$, тангенс возрастает. При этом, когда x приближается к $\frac{\pi}{2}$, значение $\sin x$ близко к единице, а значение $\cos x$ близко к нулю. Поэтому отношение $\frac{\sin x}{\cos x}$ становится сколь угодно большим. То, что любое действительное число может быть значением тангенса, видно из рис. 180.

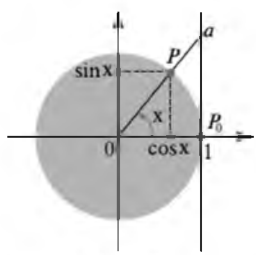


Рис. 180

Построим ось, параллельную оси ординат с началом в точке P_0. Возьмем на этой оси точку, соответствующую произвольно выбранному числу a. Соединим O с a. Получим точку P на окружности. Пусть x — число, лежащее на отрезке $\left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$ и такое, что $\cos x$, $\sin x$ — координаты P. Тогда $\operatorname{tg} x = \frac{\sin x}{\cos x} = \frac{a}{1} = a$. Мы показали, что

областью значений тангенса является вся числовая ось $\mathbb{R}$.

8. График тангенса

На отрезке $\left(0, \frac{\pi}{2}\right)$ график тангенса можно построить по точкам, учитывая, что тангенс строго возрастает, в нуле обращается в нуль, а при приближении к $\frac{\pi}{2}$ становится сколь угодно большим (рис. 181).

Отразив построенную часть графика относительно начала координат (тангенс — нечетная функция), получим график тангенса на отрезке $\left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$ (рис. 182). Для построения полного графика разобьем числовую ось на отрезки, перенося отрезок $\left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$ вправо и влево на π, 2π, 3π и т.д.

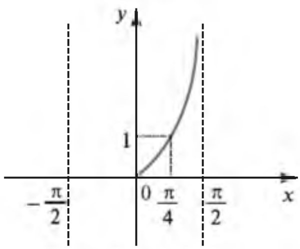


Рис. 181

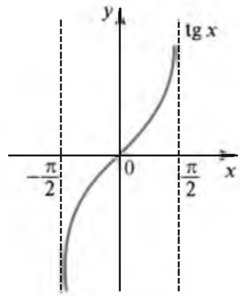


Рис. 182

График тангенса распадается на отдельные, не связанные между собой части. Это вызвано тем, что в точках $\frac{\pi}{2} + k\pi$ $(k \in \mathbb{Z})$ тангенс не определен (рис. 183).

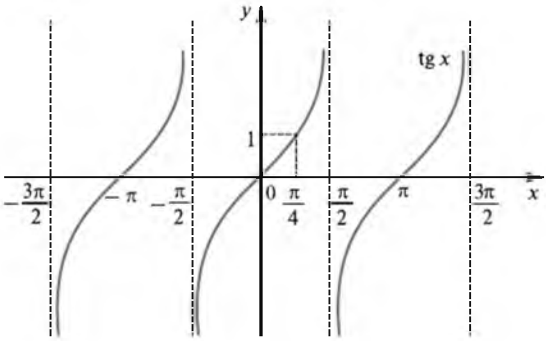


Рис. 183

Мы доказали, что тангенс возрастает на отрезке $\left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$. Можно ли сказать, что тангенс возрастает на всей области определения? Нет. Достаточно посмотреть на график. Если $x_1 = \frac{\pi}{4}$, а $x_2 = \frac{3\pi}{4}$, то $\operatorname{tg} x_1 = 1$, а $\operatorname{tg} x_2 = -1$. Хотя $x_1 < x_2$, но $\operatorname{tg} x_1 > \operatorname{tg} x_2$. Нарушение монотонности связано с тем, что между точками x_1 и x_2 лежала точка $x = \frac{\pi}{2}$, в которой тангенс не определен.

Однако верно, что тангенс возрастает на каждом отрезке, который целиком попадает в его область определения.

Котангенс

Свойства котангенса получаются так же, как и свойства тангенса. Перечислим кратко эти свойства, оставляя их доказательство для самостоятельной работы.

1. Функция $y = \operatorname{ctg} x = \dfrac{\sin x}{\cos x}$ определена при $x \neq k\pi$ $(k \in \mathbb{Z})$.

2. Функция $y = \operatorname{ctg} x$ периодична. Ее периодом является число π: $\operatorname{ctg}(x + \pi) = \operatorname{ctg} x$.

3. Функция $y = \operatorname{ctg} x$ нечетна: $\operatorname{ctg}(-x) = -\operatorname{ctg} x$.

4. Функция $y = \operatorname{ctg} x$ обращается в нуль одновременно с косинусом, т.е. для $x = \frac{\pi}{2} + k\pi$ $(k \in \mathbb{Z})$.

5. Функция $y = \operatorname{ctg} x$ положительна в первой и третьей четвертях и отрицательна во второй и четвертой.

6. Функция $y = \operatorname{ctg} x$ убывает на отрезке $(0, \pi)$. Перенося этот отрезок на $k\pi$, получаем, что котангенс убывает на каждом отрезке $(k\pi, k\pi + \pi)$ $(k \in \mathbb{Z})$.

7. Область значений котангенса — множество $\mathbb{R}$ всех действительных чисел.

8. График котангенса изображен на рис. 184.

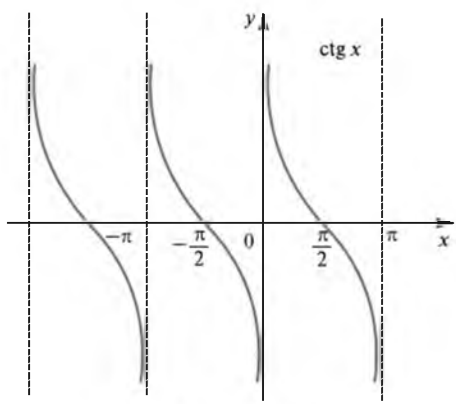


Рис. 184

Контрольные вопросы и задания

1. Как определяются синус и косинус числа t?
2. Какова область определения синуса и косинуса?
3. В каких точках синус и косинус обращаются в нуль?
4. Как меняются знаки синуса и косинуса?
5. Каковы наименьшее и наибольшее значения синуса и косинуса?
6. Являются ли точки, в которых синус (или косинус) принимает наибольшее или наименьшее значения, точками экстремума?
7. Сколько точек экстремума имеет синус (косинус)?
8. Сколько раз принимает свое наибольшее (наименьшее) значение синус (косинус)? Чему оно равно?
9. Какому неравенству удовлетворяют все значения синуса (косинуса)?
10. Каково множество значений синуса (косинуса)?
11. Можно ли сказать, что синус (косинус) является монотонной функцией?
12. На каких промежутках (в каких четвертях) синус (косинус) возрастает (убывает)?
13. Какова область определения тангенса (котангенса)?
14. Каков период тангенса (котангенса)?
15. В каких точках тангенс (котангенс) обращается в нуль?
16. Как меняются знаки тангенса (котангенса)?
17. Что можно сказать о монотонности тангенса (котангенса)?
18. Имеет ли тангенс (котангенс) точки экстремума?
19. Принимает ли тангенс (котангенс) наибольшее (наименьшее) значение?
20. Какова область значений тангенса (котангенса)?
21. Сколько раз принимает каждое свое значение тангенс (котангенс)?

§ 15. Формулы сложения

Основная формула

Тригонометрические функции связаны между собой многочисленными соотношениями. Одна их часть объясняется связью между координатами точки окружности (основные соотношения). Вторая часть тригонометрических соотношений связана с различными видами симметрии и периодичности в движении точки по окружности (формулы приведения). Третья часть тригонометрических формул, к изучению которой мы переходим, объясняется тем обстоятельством, что поворот точки на угол $\alpha + \beta$ можно составить из двух поворотов: на угол α и на угол β. Простые формулы, связывающие координаты точек $P\alpha$, $P\beta$ и $P\alpha_{+}\beta$, называются ***формулами сложения***.

Выведем формулы, выражающие $\sin(\alpha \pm \beta)$, $\cos(\alpha \pm \beta)$, $\mathrm{tg}(\alpha \pm \beta)$, $\mathrm{ctg}(\alpha \pm \beta)$ через тригонометрические функции углов α и β. Оказывается, что достаточно вывести формулу для косинуса разности. Остальные формулы получатся как ее следствия.

Теорема. Косинус разности двух углов равен произведению косинусов этих углов, сложенному с произведением синусов:

$$\cos(\alpha - \beta) = \cos\alpha\cos\beta + \sin\alpha\sin\beta.$$

Доказательство. Построим углы α и β с помощью единичной окружности, т.е. нанесем точки $P\alpha$ и $P\beta$ такие, что их радиус-векторы $\overline{OP_\alpha}$ и $\overline{OP_\beta}$ образуют углы α и β с положительным направлением оси абсцисс. Угол между векторами $\overline{OP_\alpha}$ и $\overline{OP_\beta}$ равен $\alpha - \beta$ (рис. 185). Вычислим скалярное произведение этих векторов. По определению скалярного произведения, имеем $\overline{OP_\alpha} \cdot \overline{OP_\beta} = |\overline{OP_\alpha}| \cdot |\overline{OP_\beta}| \times \cos(\alpha - \beta) = \cos(\alpha - \beta)$, так как векторы $\overline{OP_\alpha}$ и $\overline{OP_\beta}$ имеют единичную длину. Теперь вычислим это же скалярное произведение с помощью координат: $\overline{OP_\alpha} \times \overline{OP_\beta} = \cos\alpha\cos\beta + \sin\alpha\sin\beta$. Сравнивая результаты вычислений, получаем

$$\cos(\alpha - \beta) = \cos\alpha\cos\beta + \sin\alpha\sin\beta.$$

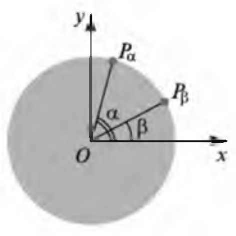


Рис. 185

Косинус суммы

Сумму $\alpha + \beta$ запишем как разность $\alpha - (-\beta)$ и подставим в формулу для косинуса разности:

$$\cos(\alpha + \beta) = \cos(\alpha - (-\beta)) = \cos\alpha\cos(-\beta) + + \sin\alpha\sin(-\beta).$$

Воспользуемся тем, что $\cos(-\beta) = \cos\beta$ (четность косинуса), а $\sin(-\beta) = -\sin\beta$ (нечетность синуса). Тогда

$$\cos(\alpha + \beta) = \cos\alpha\cos\beta - \sin\alpha\sin\beta.$$

Синус суммы

Воспользуемся одной из формул приведения:

$$\sin(\alpha+\beta)=\cos\left(\frac{\pi}{2}-(\alpha+\beta)\right)=\cos\left(\left(\frac{\pi}{2}-\alpha\right)-\beta\right).$$

Теперь используем формулу для косинуса разности:

$$\cos\left(\left(\frac{\pi}{2}-\alpha\right)-\beta\right)=\cos\left(\frac{\pi}{2}-\alpha\right)\cos\beta+\sin\left(\frac{\pi}{2}-\alpha\right)\sin\beta.$$

Окончательно

$$\sin(\alpha+\beta)=\sin\alpha\cos\beta+\cos\alpha\sin\beta.$$

Синус разности

$$\sin(\alpha-\beta)=\sin(\alpha+(-\beta))=\sin\alpha\cos(-\beta)+\cos\alpha\sin(-\beta)=$$
$$=\sin\alpha\cos\beta-\cos\alpha\sin\beta.$$

или окончательно

$$\sin(\alpha-\beta)=\sin\alpha\cos\beta-\cos\alpha\sin\beta.$$

Пример

Вычислим $\sin 15°$.

Представим $15°$ как разность $45° - 30°$. Тогда $\sin 15° = \sin(45° - 30°) =$
$= \sin 45° \cdot \cos 30° - \cos 45° \cdot \sin 30° = \frac{\sqrt{2}}{2}\cdot\frac{\sqrt{3}}{2}-\frac{\sqrt{2}}{2}\cdot\frac{1}{2}=\frac{\sqrt{2}}{4}(\sqrt{3}-1)$.

Тангенс суммы и разности

По определению, $\operatorname{tg}(\alpha+\beta)=\frac{\sin(\alpha+\beta)}{\cos(\alpha+\beta)}$.

По формулам синуса и косинуса суммы имеем

$$\operatorname{tg}(\alpha+\beta)=\frac{\sin\alpha\cos\beta+\cos\alpha\sin\beta}{\cos\alpha\cos\beta-\sin\alpha\sin\beta}.$$

Разделив числитель и знаменатель этой дроби на $\cos\alpha\cos\beta$, найдем

$$\operatorname{tg}(\alpha+\beta)=\frac{\operatorname{tg}\alpha+\operatorname{tg}\beta}{1-\operatorname{tg}\alpha\operatorname{tg}\beta}.$$

Заменяя β на $-\beta$ и пользуясь нечетностью тангенса, получаем

$$\operatorname{tg}(\alpha-\beta)=\frac{\operatorname{tg}\alpha-\operatorname{tg}\beta}{1+\operatorname{tg}\alpha\operatorname{tg}\beta}.$$

Тригонометрические функции двойных углов

Формулы сложения являются одними из основных формул, связывающих тригонометрические функции. Из них можно вывести различные следствия. Полагая $\alpha=\beta$, получим *тригонометрические функции двойных углов*:

1. $\sin 2\alpha=\sin(\alpha+\alpha)=\sin\alpha\cos\alpha+\cos\alpha\sin\alpha=2\sin\alpha\cos\alpha$;

$$\sin 2\alpha=2\sin\alpha\cos\alpha.$$

2. $\cos 2\alpha=\cos(\alpha+\alpha)=\cos\alpha\cos\alpha-\sin\alpha\sin\alpha=\cos^2\alpha-\sin^2\alpha$;

$$\cos 2\alpha=\cos^2\alpha-\sin^2\alpha.$$

Заметим, что в формуле для $\cos 2\alpha$ можно заменить $\cos^2\alpha$ на $1-\sin^2\alpha$ или $\sin^2\alpha$ на $1-\cos^2\alpha$. Получим две новые формулы:

$$\cos 2\alpha=1-2\sin^2\alpha;$$
$$\cos 2\alpha=2\cos^2\alpha-1.$$

3. $$\operatorname{tg}2\alpha=\operatorname{tg}(\alpha+\alpha)=\frac{\operatorname{tg}\alpha+\operatorname{tg}\beta}{1-\operatorname{tg}\alpha\operatorname{tg}\beta}=\frac{2\operatorname{tg}\alpha}{1-\operatorname{tg}^2\alpha};$$

$$\operatorname{tg}2\alpha=\frac{2\operatorname{tg}\alpha}{1-\operatorname{tg}^2\alpha}.$$

Тригонометрические функции половинного угла

Из формул двойных углов $\cos 2\alpha=2\cos^2\alpha-1=1-2\sin^2\alpha$ можно получить формулы для синуса и косинуса половинного угла. Сначала запишем

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

Затем в этих формулах подставим $\frac{\alpha}{2}$ вместо α:

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

$$\cos\frac{\alpha}{2}=\pm\sqrt{\frac{1+\cos\alpha}{2}},\ \sin\frac{\alpha}{2}=\pm\sqrt{\frac{1-\cos\alpha}{2}}.$$

Для того чтобы правильно выбрать знак перед корнем, надо определить, в какой четверти лежит угол $\frac{\alpha}{2}$.

Тангенс половинного угла

Обилие тригонометрических формул связано с тем, что между основными тригонометрическими функциями — синусом, косинусом, тангенсом и котангенсом — есть соотношения, которые позволяют по-разному записать одно и то же выражение. Так, $\cos 2\alpha$ можно выразить через $\cos\alpha$ и $\sin\alpha$: $\cos 2\alpha = \cos^2\alpha - \sin^2\alpha$, а можно только через $\cos\alpha$ или только через $\sin\alpha$: $\cos 2\alpha = 2\cos^2\alpha - 1$, $\cos 2\alpha = 1 - 2\sin^2\alpha$. Нетрудно при желании выразить $\cos 2\alpha$ через $\operatorname{tg}\alpha$ (эту формулу мы получим далее).

Нельзя ли выбрать одну какую-то функцию и через нее выражать все остальные? Если в качестве такой функции взять синус, то во многих формулах появятся квадратные корни. Так, выражая $\sin 2\alpha$ через $\sin\alpha$, находим

$$\sin 2\alpha = 2\sin\alpha\cos\alpha = 2\sin\alpha\left(\pm\sqrt{1-\sin^2\alpha}\right).$$

Такие формулы неудобны. Оказывается, что все тригонометрические функции от аргумента (и от кратных этого аргумента) выражаются через тангенс половинного угла рационально, без квадратных корней. Выведем эти полезные формулы.

Напишем *формулы двойного угла для исходного угла* $\frac{x}{2}$:

$$\sin x = 2\sin\frac{x}{2}\cos\frac{x}{2}, \quad \cos x = \cos^2\frac{x}{2} - \sin^2\frac{x}{2}.$$

Представим число 1 в виде $1 = \sin^2\frac{x}{2} + \cos^2\frac{x}{2}$ и разделим на 1 правые части последних формул:

$$\sin x = \frac{2\sin\left(\frac{x}{2}\right)\cos\left(\frac{x}{2}\right)}{\sin^2\left(\frac{x}{2}\right)+\cos^2\left(\frac{x}{2}\right)},$$

$$\cos x = \frac{\cos^2\left(\frac{x}{2}\right)-\sin^2\left(\frac{x}{2}\right)}{\sin^2\left(\frac{x}{2}\right)+\cos^2\left(\frac{x}{2}\right)}.$$

Разделим числитель и знаменатель каждой дроби на $\cos^2\frac{x}{2}$ и заменим $\frac{\sin\left(\frac{x}{2}\right)}{\cos\left(\frac{x}{2}\right)}$ на $\operatorname{tg}\frac{x}{2}$:

$$\sin x = \frac{2\operatorname{tg}\left(\frac{x}{2}\right)}{1+\operatorname{tg}^2\left(\frac{x}{2}\right)},\quad \cos x = \frac{1-\operatorname{tg}^2\left(\frac{x}{2}\right)}{1+\operatorname{tg}^2\left(\frac{x}{2}\right)}.$$

Пользуясь этими формулами, функцию вида $y = a\sin x + b\cos x + c$ можно представить в виде рациональной функции от $\operatorname{tg}\frac{x}{2}$.

Пример

Преобразовать функцию $y = 2\sin x + 3\cos x - 1$ к функции от $\operatorname{tg}\frac{x}{2}$:

$$2\sin x + 3\cos x - 1 = 2\frac{2\operatorname{tg}\frac{x}{2}}{1+\operatorname{tg}^2\frac{x}{2}} + 3\frac{1-\operatorname{tg}^2\frac{x}{2}}{1+\operatorname{tg}^2\frac{x}{2}} - 1 =$$

$$= \frac{4\operatorname{tg}\frac{x}{2} + 3 - 3\operatorname{tg}^2\frac{x}{2} - 1 - \operatorname{tg}^2\frac{x}{2}}{1+\operatorname{tg}^2\frac{x}{2}} =$$

$$= \frac{-4\operatorname{tg}\frac{x}{2} + 4\operatorname{tg}\frac{x}{2} + 2}{1+\operatorname{tg}^2\frac{x}{2}} = (-2)\,\frac{2\operatorname{tg}^2\frac{x}{2} - 2\operatorname{tg}\frac{x}{2} - 1}{1+\operatorname{tg}^2\frac{x}{2}}.$$

Преобразование суммы тригонометрических функций в произведение

Преобразуем сумму $\sin\alpha + \sin\beta$. Введем следующий искусственный прием. Напишем тождества $\alpha = \frac{\alpha+\beta}{2} + \frac{\alpha-\beta}{2}$ и $\beta = \frac{\alpha+\beta}{2} - \frac{\alpha-\beta}{2}$, заменим α и β выражениями, стоящими справа, в формулах для синуса суммы и разности:

$$\sin\alpha + \sin\beta = \sin\left(\frac{\alpha+\beta}{2} + \frac{\alpha-\beta}{2}\right) + \sin\left(\frac{\alpha+\beta}{2} - \frac{\alpha-\beta}{2}\right) = \sin\frac{\alpha+\beta}{2}\,\times$$

$$\times \cos\frac{\alpha-\beta}{2} + \cos\frac{\alpha+\beta}{2}\sin\frac{\alpha-\beta}{2} + \sin\frac{\alpha+\beta}{2}\cos\frac{\alpha-\beta}{2} - \cos\frac{\alpha+\beta}{2}\times$$

$$\times \sin\frac{\alpha-\beta}{2} = 2\sin\frac{\alpha+\beta}{2}\cos\frac{\alpha-\beta}{2};$$

$$\sin\alpha + \sin\beta = 2\sin\frac{\alpha+\beta}{2}\cos\frac{\alpha-\beta}{2};$$

$$\sin\alpha - \sin\beta = 2\sin\frac{\alpha-\beta}{2}\cos\frac{\alpha+\beta}{2};$$

$$\cos\alpha + \cos\beta = 2\cos\frac{\alpha+\beta}{2}\cos\frac{\alpha-\beta}{2};$$

$$\cos\alpha - \cos\beta = 2\sin\frac{\alpha+\beta}{2}\sin\frac{\beta-\alpha}{2}.$$

Преобразование произведения тригонометрических функций в сумму

Выпишем формулы сложения.

1. $\sin(\alpha+\beta) = \sin\alpha\cos\beta + \cos\alpha\sin\beta$.
2. $\sin(\alpha-\beta) = \sin\alpha\cos\beta - \cos\alpha\sin\beta$.
3. $\cos(\alpha+\beta) = \cos\alpha\cos\beta - \sin\alpha\sin\beta$.
4. $\cos(\alpha-\beta) = \cos\alpha\cos\beta + \sin\alpha\sin\beta$.

Вычтем почленно равенство 3 из 4:

$$\sin\alpha\sin\beta = \frac{1}{2}(\cos(\alpha-\beta) - \cos(\alpha+\beta)).$$

Сложим равенства 3 и 4:

$$\cos\alpha\cos\beta = \frac{1}{2}(\cos(\alpha-\beta) + \cos(\alpha+\beta)).$$

Сложим два первых равенства:

$$\sin\alpha\cos\alpha = \frac{1}{2}(\sin(\alpha+\beta) + \sin(\alpha-\beta)).$$

В этом параграфе рассмотрены различные тождества, связывающие тригонометрические функции. Все их запомнить трудно, и приходится обращаться к справочникам. Однако некоторые формулы, например, для синуса и косинуса двойного угла, вообще говоря, запомнить полезно.

Преобразование суммы тригонометрических функций в произведение связано с возможностью вычислять эти суммы с помощью логарифмических таблиц. Для компьютерных вычислений такие преобразования не обязательны.

Необходимость преобразования произведения в сумму объясняется иначе. С помощью этого преобразования можно объяснить явление модуляции в радиотехнике. В математике эти формулы нужны для интегрирования произведений тригонометрических функций.

Сложение гармонических колебаний

Как вы помните, гармоническое колебание описывается функцией вида $y = A\sin(\omega t + \alpha)$.

Необходимо научиться складывать колебания. В механике это связано с тем, что на точку может действовать несколько сил, каждая из которых вызывает гармоническое колебание. В электро- и радиотехнике сложение колебаний также является одной из важнейших операций. Оказывается, что при сложении гармонических колебаний одной и той же частоты получается снова гармоническое колебание той же частоты. На математическом языке это означает, что сумма двух функций

$$y = R_1\sin(\omega t + \alpha_1),\ y = R_2\sin(\omega t + \alpha_2)$$

есть функция того же вида:

$$y = R_3\sin(\omega t + \alpha_3).$$

Достаточно научиться складывать функции вида $y = R_1\sin\omega t$ и $y = R_2\cos\omega t$. Для их сложения применяется прием введения вспомогательного угла. Итак, рассмотрим выражение $y = R_1\sin\omega t + R_2\cos\omega t$. Оно похоже на формулу синуса суммы: $\sin(\omega t + \alpha) = \sin\omega t\cos\alpha + \cos\omega t\sin\alpha$. Числа R_1 и R_2 нельзя считать косинусом и синусом, однако их можно разделить на число $\sqrt{R_1^2 + R_2^2}$ и тогда это будет возможно. Введем угол α с помощью соотношений: $\cos\alpha = \dfrac{R_1}{\sqrt{R_1^2 + R_2^2}}$, $\sin\alpha = \dfrac{R_2}{\sqrt{R_1^2 + R_2^2}}$. Так как $\left(\dfrac{R_1}{\sqrt{R_1^2 + R_2^2}}\right)^2 + \left(\dfrac{R_2}{\sqrt{R_1^2 + R_2^2}}\right)^2 = 1$, то это возможно.

Тогда

$$R_1\sin\omega t + R_2\cos\omega t = \sqrt{R_1^2 + R_2^2}\left(\frac{R_1}{\sqrt{R_1^2 + R_2^2}}\sin\omega t + \frac{R_2}{\sqrt{R_1^2 + R_2^2}}\cos\omega t\right) =$$

$$= R_3\sin(\omega t + \alpha),\ \text{где } R_3 = \sqrt{R_1^2 + R_2^2}.$$

Сумма двух колебаний с различными периодами не обязательно будет периодическим колебанием. Интересное явление происходит при сложении колебаний с различными, но очень близкими периодами. Рассмотрим, например, сумму функций $y_1 = \sin \omega_1 t$ и $y_2 = \sin \omega_2 t$, где ω_1 и ω_2 близки друг к другу. Складывая синусы, получим $y_1 + y_2 = 2\sin \frac{\omega_1 + \omega_2}{2} t \cos \frac{\omega_1 - \omega_2}{2} t$. Так как $\omega_1 \approx \omega_2$, то $\frac{\omega_1 + \omega_2}{2} \approx \omega_1$, а $\frac{\omega_1 - \omega_2}{2} \approx 0$. Поэтому $\cos \frac{\omega_1 - \omega_2}{2} t \approx 1$ при маленьких значениях t и $y_1 + y_2 \approx 2y_1$. Однако с ростом t множитель $2 \cos \frac{\omega_1 - \omega_2}{2} t$ будет убывать. «Ровное» гармоническое колебание типа y_1 заменится «биением», график которого изображен на рис. 186. Можно представить, что биение — это колебание, амплитуда которого медленно (и тоже периодически) меняется. Явление биения можно наблюдать при наложении звуков близкой частоты, при измерении значений океанских приливов, которые вызываются наложением двух периодических процессов с близкими, но различными периодами, — притяжением Солнца и притяжением Луны.

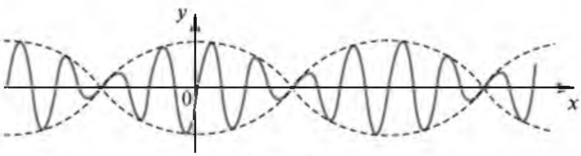


Рис. 186

Разложение на гармоники

Чистый звуковой тон представляет колебание с некоторой постоянной частотой. Музыка, которую мы слышим, — это наложение различных чистых тонов, т.е. сложение колебаний с различными частотами. Преобладание звука той или иной частоты (например, низких звуков или высоких) связано с амплитудой соответствующих колебаний. Это разложение звуков на чистые тона часто встречается при изучении различных колебательных процессов.

Можно сказать так — простейшие гармонические колебания являются теми кирпичиками, из которых можно сложить любое колебание. На языке математики это означает, что любую периодическую функцию можно представить с наперед заданной точностью как сумму синусов.

Этот замечательный факт был обнаружен еще в XVIII в. Д. Бернулли при решении задачи о колебании струны. Это показалось удивительным и невозможным по отношению к любой функции даже такому гениальному математику, как Л. Эйлер (автор всей современной символики тригонометрии). Систематически разложения периодических функций в сумму синусов (или, как говорят, на гармоники) изучал в начале XIX в. французский математик Ж. Фурье. Их так теперь и называют — *разложения* (или *ряды*) *Фурье*.

В качестве примера на рис. 187 изображено приближение к периодической функции $y = \{x\}$ в виде суммы нескольких гармоник. Разложение произвольного периодического сигнала на гармоники является главным математическим аппаратом радиотехники.

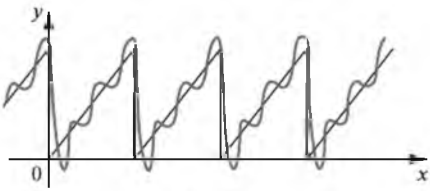


Рис. 187

Дробную часть числа x обозначают $\{x\}$. Например, $\{1{,}7\} = 0{,}7$; $\{15\} = 0$; $\{-3{,}2\} = 0{,}8$. Функция $y = \{x\}$ является периодической с периодом $T = 1$ (рис. 188).

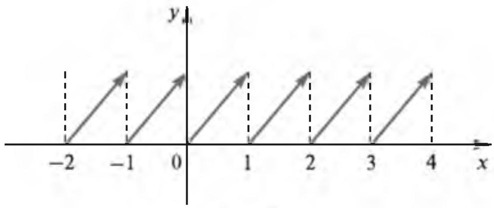


Рис. 188

Контрольные вопросы и задания

1. Чему равен косинус разности двух углов?
2. Как, зная формулу для косинуса разности, вывести другие формулы сложения?
3. Приведите все известные вам формулы для косинуса двойного угла.

4. Через какую функцию все тригонометрические функции выражаются с помощью рациональных формул?
5. Какой искусственный прием применяется для вывода формул для суммы и разности синусов и косинусов?
6. Какой формулой можно задать гармоническое колебание?
7. Чему равна амплитуда суммы двух гармонических колебаний одной и той же частоты?
8. Какая функция называется периодической?
9. Приведите примеры периодических функций с периодом $T = 1$.
10. Что можно сказать о периодичности суммы периодических функций?

§ 16. Тригонометрические уравнения

Уравнение $\sin x = a$

Простейшими тригонометрическими уравнениями считают уравнения вида $\sin x = a$, $\cos x = a$, $\operatorname{tg} x = a$, $\operatorname{ctg} x = a$.

Как уже известно, область значений синуса — отрезок $[-1, 1]$. Если $|a| > 1$, то уравнение $\sin x = a$ решений не имеет. Пусть теперь $|a| \leq 1$. Построим на одном рисунке графики $y = a$ и $y = \sin x$.

На рисунке 189 видно, что прямая $y = a$ пересекает синусоиду бесконечно много раз. Это означает, что при $|a| \leq 1$ уравнение $\sin x = a$ имеет бесконечно много корней. Так как синус имеет период 2π, то достаточно найти все решения в пределах одного периода. По графику (см. рис. 189) видно, что при $|a| < 1$ на отрезке $[0, 2\pi]$ есть два числа, синус которых равен a.

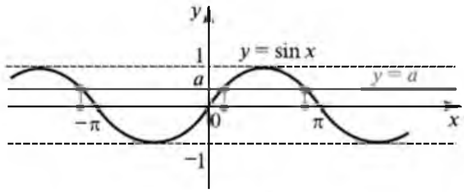


Рис. 189

То же самое можно получить, рассматривая движение точки по окружности: в пределах одного оборота найдутся два угла с одинаковыми значениями синуса.

Если один из таких углов α, то другой $\pi - \alpha$. Все другие решения уравнения $\sin x = a$ ($|a| < 1$) получаются из двух найденных прибавлением периода.

Итак, пусть α — какое-либо решение уравнения $\sin x = a$ ($|a| < 1$). Тогда все решения этого уравнения получаются по формулам (рис. 190).

$$x = \alpha + 2k\pi,\ x = \pi - \alpha + 2k\pi\ (k \in \mathbb{Z}).$$

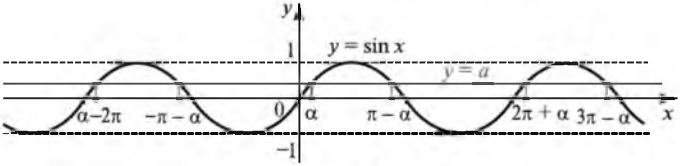


Рис. 190

Эти две серии решений указаны на графике и иногда записываются одной формулой

$$x = (-1)^n\alpha + \pi n\ (n \in \mathbb{Z}).$$

Пример

Решить уравнение $\sin x = \frac{1}{2}$.

Одно решение этого уравнения: $x = \frac{\pi}{6}$. Все остальные решения получаются по формулам

$$x = \frac{\pi}{6} + 2k\pi \text{ и } x = \pi - \frac{\pi}{6} + 2k\pi = \frac{5\pi}{6} + 2k\pi\ (k \in \mathbb{Z}).$$

Арксинус

Как мы уже выяснили, уравнение $\sin x = a$ при $|a| \leq 1$ имеет бесконечно много решений. Для одного из них имеется специальное название — арксинус.

Пусть число a по модулю не превосходит единицы. Арксинусом числа a называется угол x, лежащий на отрезке $\left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$, синус которого равен a.

Обозначение: $x = \arcsin a$.

Итак, равенство $x = \arcsin a$ равносильно двум условиям: $\sin x = a$ и $-\frac{\pi}{2} \leq x \leq \frac{\pi}{2}$. Обратим внимание на то, что $\arcsin a$ существует лишь если $|a| \leq 1$.

Примеры

1. $\arcsin 0 = 0$. Действительно, $0 \in \left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$ и $\sin 0 = 0$.
2. $\arcsin \frac{1}{2} = \frac{\pi}{6}$.
3. $\arcsin\left(-\frac{1}{2}\right) = -\frac{\pi}{6}$.

Теперь решения уравнения $\sin x = a$ (при $|a| \le 1$) можно представить формулами

$$x = \arcsin a + 2\pi k \text{ и } x = \pi - \arcsin a + 2\pi k \ (k \in \mathbb{Z})$$

или в виде одной формулы

$$x = (-1)^k \arcsin a + \pi k \ (k \in \mathbb{Z}).$$

Отметим некоторые тождества:

1. $\sin(\arcsin a) = a$. Действительно, согласно определению, $\arcsin a$ — это такой угол x, что $\sin x = a$, т.е. $\sin(\arcsin a) = a$. Здесь не использовано условие $\arcsin a \in \left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$.
2. $\arcsin(\sin x) = x$, если $x \in \left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$. Действительно, обозначим $\sin x$ через a. Тогда тождество превратится в определение арксинуса: $\arcsin a = x$, если $x \in \left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$ и $\sin x = a$. Заметим, что выражение $\arcsin(\sin x)$ имеет смысл при любом x, однако при $x \notin \left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$ оно не равно x.
3. $\arcsin(-a) = -\arcsin a$. Действительно, синусы от правой и левой частей равны $\sin(\arcsin(-a)) = -a$ и $\sin(-\arcsin a) = -\sin(\arcsin a) = -a$. В то же время правая часть — это угол, лежащий в пределах $\left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$. Поэтому левая и правая части равны между собой.

Уравнение $\cos x = a$

Так же как и в предыдущем случае, получаем, что при $|a| > 1$ уравнение $\cos x = a$ решения не имеет, при $|a| \le 1$ решений бесконечно много.

Если α — какое-либо решение уравнения $\cos x = a$, то $-\alpha$ также является решением ($\cos \alpha = \cos(-\alpha)$). По графику (рис. 191) или на тригонометрическом круге видно, что в пределах одного периода уравнение $\cos x = a$ имеет два решения при $|a| < 1$.

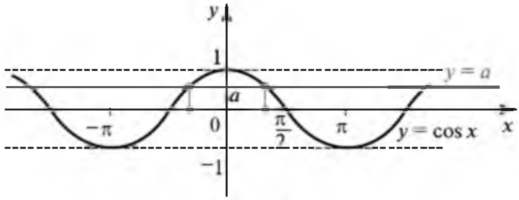


Рис. 191

Если α — одно решение уравнения $\cos x = a$, то все решения исчерпываются двумя сериями:

$$x = \alpha + 2k\pi, x = -\alpha + 2\pi k \ (k \in \mathbb{Z}).$$

Эти серии обычно записывают в виде одной формулы

$$x = \pm\alpha + 2\pi k \ (k \in \mathbb{Z}).$$

Пример

Решить уравнение $\cos x = -\frac{1}{2}$.

Одно решение находится легко: $x = \frac{2}{3}\pi$. Все остальные решения уравнения $\cos x = -\frac{1}{2}$: $x = \pm\frac{2}{3}\pi + 2k\pi \ (k \in \mathbb{Z})$.

Арккосинус

Так же как и для синуса, выделяется одно определенное решение уравнения $\cos x = a$, которое называется арккосинусом.

Пусть a — число, по модулю не превосходящее единицы. Арккосинусом числа a называется угол x, лежащий на отрезке $[0, \pi]$, косинус которого равен a.

Обозначение: $x = \arccos a$.

Равенство $x = \arccos a$ равносильно двум условиям: $\cos x = a$ и $0 \le x \le \pi$. Арккосинус числа a существует лишь при $-1 \le a \le 1$.

Примеры

$$\arccos 0 = \frac{\pi}{2}, \arccos \frac{1}{2} = \frac{\pi}{3}, \arccos\left(-\frac{1}{2}\right) = \frac{2}{3}\pi.$$

Решение уравнения $\cos x = a$ $|a| \le 1$ можно записать в общем виде: $x = \pm\arccos a + 2\pi k \ (k \in \mathbb{Z})$.

По каким причинам для значений арксинуса был выбран отрезок $\left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$, а для арккосинуса — $[0, \pi]$? Выбраны эти отрезки, так как на них, во-первых, синус и косинус принимают все возможные значения от -1 до 1 и, во-вторых, каждое значение принимается ровно один раз. Отрезков, удовлетворяющих этим условиям, бесконечное множество, но при этом выбраны отрезки, «наиболее близкие к нулю».

Для арккосинуса можно вывести ряд тождеств.

1. $\cos(\arccos a) = a$. Это часть определения арккосинуса.
2. $\arccos(\cos x) = x$, если $x \in [0, \pi]$. Обозначим $\cos x = a$. Получим определение арккосинуса: $\arccos a = x$, если $x \in [0, \pi]$ и $\cos x = a$.
3. $\arccos(-a) = \pi - \arccos a$.

Вычислим косинус от обеих частей этого равенства, используя первое тождество:

$$\cos(\arccos(-a)) = -a,$$
$$\cos(\pi - \arccos a) = -\cos(\arccos a) = -a.$$

Если равны косинусы двух чисел, то это еще не означает, что равны сами числа. Проверим, что правая часть лежит на отрезке $[0, \pi]$. Так как левая часть тоже лежит на этом отрезке, то из равенства косинусов двух чисел теперь уже будет следовать равенство самих чисел, так как на отрезке $[0, \pi]$ косинус принимает каждое значение ровно один раз.

Итак, надо доказать, что $\pi - \arccos a \in [0, \pi]$. Действительно, $\arccos a \in [0, \pi]$, $-\arccos a \in [-\pi, 0]$, $\pi - \arccos a \in [0, \pi]$, что и требовалось доказать.

Решение уравнений $\operatorname{tg} x = a$ и $\operatorname{ctg} x = a$

Область значений тангенса (котангенса) — вся числовая ось $\mathbb{R}$. Поэтому уравнения $\operatorname{tg} x = a$ и $\operatorname{ctg} x = a$ имеют решения при любом a. В пределах одного периода (а он равен π) тангенс и котангенс принимают каждое значение ровно один раз. Поэтому если известно одно решение уравнения $\operatorname{tg} x = a$ и $\operatorname{ctg} x = a$, то все остальные получаются прибавлением периода:

$$\operatorname{tg} x = a \Rightarrow x = \alpha + k\pi,\ \operatorname{ctg} x = a \Rightarrow x = \alpha + k\pi\ (k \in \mathbb{Z}),$$

где α — какое-либо решение соответствующего уравнения.

Примеры

Решите уравнения:

1. $\operatorname{tg} x = 1, x = \frac{\pi}{4} + k\pi \ (k \in \mathbb{Z})$.
2. $\operatorname{ctg} x = -\frac{\sqrt{3}}{3}, x = -\frac{\pi}{3} + k\pi \ (k \in \mathbb{Z})$.

Арктангенс и арккотангенс

Определения арктангенса и арккотангенса аналогичны определениям арксинуса и арккосинуса.

Арктангенсом числа a называется угол $x \in \left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$, тангенс которого равен a.
Арккотангенсом числа a называется угол $x \in (0, \pi)$, котангенс которого равен a.

Обозначения: $x = \operatorname{arctg} a$, $x = \operatorname{arcctg} a$.

Примеры

$$\operatorname{arctg} 0 = 0, \operatorname{arcctg} 0 = \frac{\pi}{2}, \operatorname{arctg}(-1) = -\frac{\pi}{4}, \operatorname{arcctg}(-1) = \frac{3\pi}{4}.$$

При решении уравнений функции $\operatorname{arctg} x$, $\operatorname{arcctg} x$ используются так:

$$\operatorname{tg} x = a \Leftrightarrow x = \operatorname{arctg} a + k\pi \ (k \in \mathbb{Z});$$
$$\operatorname{ctg} x = a \Leftrightarrow x = \operatorname{arcctg} a + k\pi \ (k \in \mathbb{Z}).$$

Основные тождества

1. $\operatorname{tg} \operatorname{arctg} a = a$, $\operatorname{ctg} \operatorname{arcctg} a = a$.
2. $\operatorname{arctg}(\operatorname{tg} x) = x$, если $x \in \left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$; $\operatorname{arcctg}(\operatorname{ctg} x) = x$, если $x \in (0, \pi)$.
3. $\operatorname{arctg}(-x) = -\operatorname{arctg} x$; $\operatorname{arcctg}(-x) = \pi - \operatorname{arcctg} x$.

Мы привели простейшие тригонометрические уравнения. Более сложные тригонометрические уравнения обычно решаются сведением их к простейшим с помощью различных алгебраических и тригонометрических формул и преобразований. Рассмотрим некоторые приемы решения тригонометрических уравнений.

Уравнения, алгебраические относительно одной из тригонометрических функций

Примеры

1. Решить уравнение $2\sin^2 x + 3\sin x - 2 = 0$.

   Это уравнение является квадратным относительно $\sin x$. Его корни: $\sin x = \frac{1}{2}$, $\sin x = -2$. Второе из полученных простейших уравнений не имеет решений, так как $|\sin x| \le 1$, решения первого можно записать так:

   $$x = \frac{\pi}{6} + 2k\pi,\ x = \pi - \frac{\pi}{6} + 2\pi k = \frac{5\pi}{6} + 2\pi k\ (k \in \mathbb{Z}).$$

   Если в уравнении встречаются разные тригонометрические функции, то надо пытаться их заменить на какую-нибудь одну, используя тригонометрические тождества.

2. Решить уравнение $2\sin^2 x - 5\cos x - 5 = 0$.

   Так как квадрат синуса легко выражается через косинус, то, заменяя $\sin^2 x$ на $1 - \cos^2 x$ и приводя уравнение к квадратному относительно $\cos x$, получим $2(1 - \cos^2 x) - 5\cos x - 5 = 0$, т.е. квадратное уравнение $2\cos^2 x + 5\cos x + 3 = 0$, корни которого $\cos x = -1$, $\cos x = -\frac{3}{2}$. Уравнение $\cos x = -\frac{3}{2}$ решений не имеет. Решения уравнения $\cos x = -1$ запишем в виде $x = \pi + 2k\pi$ $(k \in \mathbb{Z})$.

3. Решить уравнение $\operatorname{tg} x + 3\operatorname{ctg} x = 4$.

   Заменяя $\operatorname{ctg} x$ на $\frac{1}{\operatorname{tg} x}$ и приводя выражение к общему знаменателю, получаем квадратное уравнение $\operatorname{tg}^2 x - 4\operatorname{tg} x + 3 = 0$, корни которого $\operatorname{tg} x = 1$, $\operatorname{tg} x = 3$, откуда $x = \frac{\pi}{4} + k\pi$, $x = \operatorname{arctg} 3 + k\pi$ $(k \in \mathbb{Z})$.

4. Решить уравнение $2\sin x + \cos x = 2$.

   Если в этом уравнении заменим косинус на синус (по аналогии с предыдущими примерами) или наоборот, то получим уравнение с радикалами. Чтобы избежать этого, используем формулы, выражающие синус и косинус через тангенс половинного угла:

   $$\sin x = \frac{2\operatorname{tg}\left(\frac{x}{2}\right)}{1+\operatorname{tg}^2\left(\frac{x}{2}\right)} \text{ и } \cos x = \frac{1-\operatorname{tg}^2\left(\frac{x}{2}\right)}{1+\operatorname{tg}^2\left(\frac{x}{2}\right)}.$$

Делая замену, получаем уравнение относительно $\operatorname{tg}\frac{x}{2}$:

$$4\operatorname{tg}\frac{x}{2}+1-\operatorname{tg}^2\frac{x}{2}=2\left(1+\operatorname{tg}^2\frac{x}{2}\right).$$

Квадратное уравнение $3\operatorname{tg}^2\frac{x}{2}-4\operatorname{tg}\frac{x}{2}+1=0$ имеет корни $\operatorname{tg}\frac{x}{2}=1$, $\operatorname{tg}\frac{x}{2}=\frac{1}{3}$, откуда $\frac{x}{2}=\frac{\pi}{4}+k\pi \Leftrightarrow x=\frac{\pi}{2}+2k\pi$; $\frac{x}{2}=\operatorname{arctg}\frac{1}{3}+k\pi \Leftrightarrow$ $x=2\operatorname{arctg}\frac{1}{3}+2\pi k$ $(k\in\mathbb{Z})$.

Это же уравнение можно решить другим способом, вводя вспомогательный угол:

$$2\sin x+\cos x=\sqrt{5}\left(\frac{2}{\sqrt{5}}\sin x+\frac{1}{\sqrt{5}}\cos x\right).$$

Пусть $\alpha=\arccos\frac{2}{\sqrt{5}}$. Тогда можно продолжить преобразование: $2\sin x+\cos x=\sqrt{5}\sin(x+\alpha)$. Получаем простейшее уравнение $\sqrt{5}\sin(x+\alpha)=2$, т.е. $\sin(x+\alpha)=\frac{2}{\sqrt{5}}$, откуда $x+\alpha=$ $=\arcsin\frac{2}{\sqrt{5}}+2k\pi$, или $x+\alpha=\pi-\arcsin\frac{2}{\sqrt{5}}+2k\pi$.

Ответ получился в другом виде, однако можно проверить, что решения на самом деле совпадают.

При замене синуса и косинуса на тангенс половинного угла может произойти потеря корней. Это связано с тем, что при $x=\pi$ не существует $\operatorname{tg}\frac{x}{2}$. Поэтому при использовании указанной замены надо проверять, не являются ли углы x, при которых $\cos x=-1$ ($a\sin x=0$), решениями уравнения.

Пример

$3\sin x-4\cos x=4$.

Переход к $\operatorname{tg}\frac{x}{2}$ дает уравнение $\operatorname{tg}\frac{x}{2}=\frac{4}{3}$, но при этом теряются решения уравнения $\cos x=-1$, которые тоже являются корнями исходного уравнения.

Понижение порядка уравнения

Формулы удвоения

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

позволяют квадраты синуса, косинуса и их произведения заменять линейными функциями от синуса и косинуса двойного угла. Такие замены делать выгодно, так как они понижают порядок уравнения.

Примеры

1. Решить уравнение $\cos 2x + \cos^2 x = \frac{5}{4}$.

   Можно заменить $\cos 2x$ на $2\cos^2 x - 1$ и найти квадратное уравнение относительно $\cos x$, но проще заменить $\cos^2 x$ на $\frac{1}{2}(1 + \cos 2x)$ и получить линейное уравнение относительно $\cos 2x$.

2. Решить уравнение $\sin^4 x + \cos^4 x = \frac{65}{81}$.

   Подставляя вместо $\sin^2 x$, $\cos^2 x$ их выражения через $\cos 2x$, получаем:

$$\frac{1}{4}(1 - \cos 2x)^2 + \frac{1}{4}(1 + \cos 2x)^2 = \frac{65}{81};$$

$$1 - 2\cos 2x + \cos^2 2x + 1 + 2\cos 2x + \cos^2 2x = 4\cdot\frac{65}{81};$$

$$\cos^2 2x = 2\cdot\frac{65}{81} - 1,\ \cos^2 2x = \frac{49}{81},\ \cos 2x = \pm\frac{7}{9};$$

$$2x = \pm\arccos\frac{7}{9} + 2k\pi,\ \ x = \pm\frac{1}{2}\arccos\frac{7}{9} + \frac{k\pi}{2}\ (k\in\mathbb{Z}).$$

Использование тригонометрических формул сложения и следствий из них

Иногда в уравнениях встречаются тригонометрические функции кратных углов. В таких случаях нужно использовать формулы сложения.

Примеры

1. Решить уравнение $\sin x + \sin 2x + \sin 3x = 0$.
   Сложим два крайних слагаемых: $(\sin x + \sin 3x) + \sin 2x = 0$, откуда $2\sin 2x\cos x + \sin 2x = 0$, $\sin 2x(2\cos x + 1) = 0$. Тогда $\sin 2x = 0$, $2x = k\pi$, $x = \frac{1}{2}k\pi$ или $2\cos x = -1$, $x = \pm\frac{2}{3}\pi + 2k\pi$ $(k \in \mathbb{Z})$.
2. Решить уравнение $\sin 3x \sin 5x = \sin x \sin 7x$.
   Преобразуем произведение синусов в сумму:
   $\frac{1}{2}(\cos 2x - \cos 8x) = \frac{1}{2}(\cos 6x - \cos 8x)$,
   откуда $\cos 2x = \cos 6x$. Полученное уравнение можно решить разными способами: 1) использовать формулы сложения; 2) преобразовать $\cos 6x - \cos 2x$ в произведение. Удобнее воспользоваться условием равенства косинусов двух углов $2x$ и $6x$: $6x = \pm 2x + 2\pi k$ $(k \in \mathbb{Z})$. Получаем два уравнения:

$$6x = 2x + 2\pi k,\ 4x = 2\pi k,\ x = \frac{1}{2}k\pi,$$

$$6x = -2x + 2k\pi,\ 8x = 2k\pi,\ x = \frac{1}{4}k\pi.$$

   Проверьте, что решения второй серии содержат в себе все решения первой серии. Учитывая это, ответ можно записать короче: $x = \frac{k\pi}{4}$ $(k \in \mathbb{Z})$.

Однородные уравнения

Рассмотрим уравнение $\sin^2 x - 5\sin x\cos x + 6\cos^2 x = 0$. Если считать, что $\sin x$ и $\cos x$ — члены первой степени, то каждое слагаемое имеет степень 2.

Уравнение, в котором каждое слагаемое имеет одну и ту же степень, называется однородным.

Его можно решить, выполнив деление на старшую степень синуса (или косинуса). Делим наше уравнение на $\cos^2 x$. Заметим: если в данное уравнение подставить $\cos x = 0$, то получим $\sin x = 0$, что невозможно, значит, в результате деления на $\cos^2 x$ не будет потери корней. Продолжая решение, находим:

$$\frac{\sin^2 x}{\cos^2 x} - 5\frac{\sin x}{\cos x} + 6 = 0,\ \operatorname{tg}^2 x - 5\operatorname{tg} x + 6 = 0,\ \operatorname{tg} x = 2,\ \operatorname{tg} x = 3,$$

$$x = \operatorname{arctg} 2 + k\pi,\ x = \operatorname{arctg} 3 + k\pi\ (k \in \mathbb{Z}).$$

Так как $1 = \sin^2 x + \cos^2 x$, то постоянные слагаемые можно считать членами второй степени.

Пример

$5 \sin^2 x + 3 \sin x \cos x = 4$.

Заменяя 4 на $4(\sin^2 x + \cos^2 x)$, получаем: $\sin^2 x + 3 \sin x \cos x - 4 \cos^2 x = 0$, $\operatorname{tg}^2 x + 3 \operatorname{tg} x - 4 = 0$, $\operatorname{tg} x = 1$, $\operatorname{tg} x = -4$, $x = \frac{\pi}{4} + k\pi$, $x = -\operatorname{arctg} 4 + k\pi$ $(k \in \mathbb{Z})$.

Контрольные вопросы и задания

1. Обратите внимание на следующие ключевые слова и выражения, появившиеся в параграфе: арксинус, арккосинус, арктангенс, однородное уравнение. Приведите примеры их использования.
2. Что такое $\arcsin a$?
3. Какие тождества для арксинуса вы знаете?
4. При каких a определен $\arcsin a$?
5. Какие значения принимает $\arcsin a$?
6. Сформулируйте вопросы, аналогичные 2—5 для $\arccos a$ и $\operatorname{arctg} a$, и дайте на них ответы.
7. Сколько решений имеют простейшие уравнения типа $\sin x = a$, $\cos x = a$, $\operatorname{tg} x = a$, $\operatorname{ctg} x = a$?
8. Как, зная одно из решений простейшего тригонометрического уравнения, найти все его решения?
9. Какими формулами выгоднее пользоваться при решении тригонометрических уравнений и почему?
10. Придумайте несколько различных способов решения уравнения $\sin x + \cos x = 1$.

ГЛАВА 4

КОРНИ, СТЕПЕНИ И ЛОГАРИФМЫ

§ 17. Преобразование выражений, содержащих корни, степени и логарифмы

Корни

Извлечение корня — это операция, обратная операции возведения в степень. Корнем *n*-й степени из числа *a* называется такое число *b*, *n*-я степень которого равна *a*: *b* — корень *n*-й степени из *a* $\Leftrightarrow$ *bn* = *a*.

Примеры

1. Так как $2^6 = 64$, то число 2 является корнем 6-й степени из числа 64.
2. Равенства $(-3)^3 = -27$, $(-5)^4 = 625$ можно прочесть так: число -3 является кубическим корнем из -27, -5 является корнем четвертой степени из 625.

Если число n четно, то равенство $b^n = a$ возможно лишь при $a \geq 0$, так как любое число в четной степени положительно. Это означает, что корней четной степени из отрицательных чисел не существует.

Если $b^n = a$, то $(-b)^n = a$, так как четная степень «минус единицы» равна «плюс единице». Это означает, что если b — корень четной степени из числа a, то и $(-b)$ есть корень той же степени из числа a.

Если n нечетно, то корень n-й степени можно извлечь из любого числа. В равенстве $b^n = a$ при нечетном n числа a и b имеют одинаковые знаки, поэтому корень нечетной степени из положительного числа положителен, а из отрицательного — отрицателен.

При нечетном n корень n-й степени из числа a извлекается единственным способом. Он и обозначается с помощью радикала:

$$b^n = a \Leftrightarrow b = \sqrt[n]{a}.$$

Можно записать: $\sqrt[3]{-27}=-3$, $\sqrt[5]{32}=2$, $\sqrt[17]{-1}=-1$, $\sqrt[7]{0}=0$. При четном n корень n-й степени можно извлечь только из неотрицательного числа, но при этом есть два корня n-й степени из положительного числа a: положительный и отрицательный. С помощью радикала четной степени обознаются положительный корень.

Примеры

$\sqrt[4]{16}=2$, $\sqrt[6]{1}=1$, $\sqrt{(2-\sqrt{5})^2}=\sqrt{5}-2$, $\sqrt[8]{0}=0$.

Итак, корней четной степени из положительного числа a существует два: один (положительный) $\sqrt[n]{a}$, а другой (отрицательный) $-\sqrt[n]{a}$.

Положительный корень n-й степени из положительного числа a, т.е. $\sqrt[n]{a}$, называют ***арифметическим корнем n-й степени***.

Свойства радикалов:

1. $(\sqrt[n]{a})^n=a$;
2. $\sqrt[n]{ab}=\sqrt[n]{a}\sqrt[n]{b}$;
3. $\sqrt[nk]{a^k}=\sqrt[n]{a}$;
4. $\sqrt[n]{\dfrac{a}{b}}=\dfrac{\sqrt[n]{a}}{\sqrt[n]{b}}$;
5. $(\sqrt[n]{a})^k=\sqrt[n]{a^k}$;
6. $\sqrt[n]{\sqrt[k]{a}}=\sqrt[nk]{a}$.

Если показатель n радикала четный, то считается, что под знаком радикала стоит положительное число. Доказательство всех свойств радикалов проводится по одной и той же схеме.

Примеры

1. $\sqrt[n]{ab}=\sqrt[n]{a}\sqrt[n]{b}$. Вспомним, что такое $\sqrt[n]{ab}$. По определению, это такое число, n-я степень которого равна числу ab, причем если n четно, то это число положительно. Проверим, что этим свойством обладает правая часть. Возведем в степень число $\sqrt[n]{a}\sqrt[n]{b}$, пользуясь известным свойством степени: $(\sqrt[n]{a}\sqrt[n]{b})^n=(\sqrt[n]{a})^n(\sqrt[n]{b})^n=$ $=ab$. При четном n каждый множитель $\sqrt[n]{a}$ и $\sqrt[n]{b}$ положителен, поэтому и $\sqrt[n]{a}\sqrt[n]{b}$ положительно.
2. $\sqrt[n]{\sqrt[k]{a}}=\sqrt[nk]{a}$. Возведем число $\sqrt[n]{\sqrt[k]{a}}$ в степень nk, пользуясь известным свойством степени: $(\sqrt[n]{\sqrt[k]{a}})^{nk}=\left((\sqrt[n]{\sqrt[k]{a}})^n\right)^k=(\sqrt[k]{a})^k=a$.

   Полученное равенство означает, что число $\sqrt[n]{\sqrt[k]{a}}$ является корнем степени nk из числа a. Если число nk четно, то, по условию, число a положительно, а тогда являются положительными числа $\sqrt[k]{a}$ и $\sqrt[n]{\sqrt[k]{a}}$.

Преобразования радикалов приходится делать над выражениями, которые содержат отрицательные числа. При этом надо соблюдать осторожность, в частности следить за знаками перед радикалами четной степени.

Примеры

1. Выражение $\sqrt[6]{a^2}$ имеет смысл при любом значении a, так как под корнем стоит число $a^2 \geq 0$. Однако при сокращении показателей надо учесть знак a: если $a \geq 0$, то $\sqrt[6]{a^2} = \sqrt[3]{a}$, если же $a < 0$, то $\sqrt[6]{a^2} = -\sqrt[3]{a}$. В частности, $\sqrt[6]{(3-\pi)^2} = \sqrt[3]{3-\pi}$ (или $-\sqrt[3]{3-\pi}$).
2. Выражение $\sqrt[n]{a^n}$ при четном n имеет смысл при любом значении a. При этом если $a \geq 0$, то $\sqrt[n]{a^n} = a$, если же $a < 0$, то $\sqrt[n]{a^n} = -a$. Можно использовать знак модуля и записать (при четном n)

$$\sqrt[n]{a^n} = |a|.$$

Извлечение корня n-й степени всегда считалось трудной вычислительной задачей — уже алгоритм извлечения квадратного корня вручную достаточно сложен и плохо запоминается. На современных микрокалькуляторах обычно есть операция извлечения квадратного корня. Корни более высокой степени вычисляются как значения показательной функции $y = a^x$ при рациональных значениях x.

Степени

Наша задача — определить степень a^b при любом значении b. Будем считать, что основание степени — число $a > 0$.

Понятие степени строится постепенно. Сначала надо вспомнить, что такое степень с натуральным показателем, т.е. рассмотреть случай, когда b — натуральное число. Запись 2^{10} мы понимаем как произведение десяти одинаковых множителей, каждый из которых равен 2. В общем виде, если $b = n$ — натуральное число, запись a^n означает произведение n множителей, каждый из которых равен a:

$$\underbrace{a^n = aa\ldots a}_{n\text{ раз}}.$$

Если b — отрицательное целое число, то его можно записать в виде $b = -n$, где n — натуральное число. Тогда

$$a^{-n} = \frac{1}{a^n}.$$

Если $b = 0$, то a^b принимается равным единице:

$$a^0 = 1.$$

Итак, мы определили степень с произвольным целым показателем. Рассмотрим случай, когда b — рациональное число. Его можно записать в виде дроби $b = \frac{m}{n}$, где m — целое число, n — натуральное число. Определим степень a^b с помощью корня:

$$a^{m/n} = \sqrt[n]{a^m}.$$

Так, $3^{2/3} = \sqrt[3]{3^2} = \sqrt[3]{9}$; $\left(\frac{2}{5}\right)^{-1/3} = \sqrt[3]{\left(\frac{2}{5}\right)^{-1}} = \sqrt[3]{\frac{5}{2}}$.

Степень с произвольным вещественным показателем b определяется следующим образом. Для числа b выбирается последовательность рациональных чисел $b_1, b_2, \ldots, b_n, \ldots$, задающая приближение числа b с любой степенью точности. Строится последовательность степеней с рациональными показателями: a^{b_1}, a^{b_2}, ..., a^{b_n}, ... Оказывается, эта последовательность задает приближение некоторого числа c с любой степенью точности. Это число c называется *степенью* a^b.

Таким образом, мы определили степень a^b положительного основания a с любым показателем b.

1. Степень числа с натуральным показателем имеет смысл не только для положительного, но и для любого основания, так как эта степень определяется с помощью операции умножения, а умножать можно любые числа. Поэтому имеют смысл равенства $(-3)^3 = (-3) \cdot (-3) \cdot (-3) = -27$, $(-1)^{100} = 1$, $\left(-\frac{1}{2}\right)^5 = -\frac{1}{32}$, $0^{11} = 0$ и др. Степень с целым отрицательным показателем может быть определена для любого числа, кроме нуля, так как ее вычисление сводится к операциям умножения и деления. Определение же степени с рациональным показателем требует операции извлечения корня, которая выполнима, как правило, только для положительных чисел. Поэтому мы с самого начала считаем основание положительным числом.
2. Определение степени с рациональным показателем основано на понятии корня n-й степени из числа: $a^{m/n} = \sqrt[n]{a^m}$. Из-за того, что одно и то же рациональное число может быть записано в виде различных дробей (например, $\frac{2}{3} = \frac{4}{6} = \frac{6}{9}$ и т.д.), может показать-

ся, что будет различным определение степени ($a^{2/3} = \sqrt[3]{a^2}$, $a^{4/6} = \sqrt[6]{a^4}$ и т.д.). Однако по свойству корней, $\sqrt[nk]{a^{mk}} = \sqrt[n]{a^m}$, поэтому на самом деле определение степени с рациональным показателем не зависит от того, в виде какой дроби записан этот показатель.

3. Вычисление степени с иррациональным показателем делается приближенно. Сначала мы задаем приближения к числу b с помощью рациональных чисел, затем вычисляем степени с рациональным показателем. У нас остался невыясненным вопрос: зная погрешность приближения числа b с помощью рационального числа b_k, как оценить погрешность приближения a^{b_k} к числу a^b?

Свойства степеней

1. $a^{b_1}a^{b_2} = a^{b_1+b_2}$.
2. $\dfrac{a^{b_1}}{a^{b_2}} = a^{b_1-b_2}$.
3. $(a^{b_1})^{b_2} = a^{b_1b_2}$.

Для рациональных значений показателя эти свойства сводятся к свойствам радикалов. Докажем, например, первое свойство для рациональных положительных показателей.

Пусть $b_1 = \dfrac{m_1}{n_1}$, $b_2 = \dfrac{m_2}{n_2}$. Приведем эти дроби к общему знаменателю, а затем их сложим:

$$b_1 = \frac{m_1n_2}{n_1n_2},\ b_2 = \frac{m_2n_1}{n_2n_1},\ b_1 + b_2 = \frac{m_1n_2 + m_2n_1}{n_1n_2}.$$

Тогда

$$a^{b_1}a^{b_2} = a^{\frac{m_1n_2}{n_1n_2}}a^{\frac{m_2n_1}{n_2n_1}} = \sqrt[n_1n_2]{a^{m_1n_2}}\sqrt[n_1n_2]{a^{m_2n_1}} = \sqrt[n_1n_2]{a^{m_1n_2}a^{m_2n_1}} =$$

$$= \sqrt[n_1n_2]{a^{m_1n_2+m_2n_1}} = a^{\frac{m_1n_2+m_2n_1}{n_1n_2}} = a^{b_1+b_2}.$$

Логарифмы

Пусть числа a, b и c связаны соотношением $a^b = c$. Можно сказать, что число c является степенью числа a (основания) с показателем b. Если числа a и c фиксированы, а нужно выразить через них b, то используют понятие логарифма. Будем считать, что в качестве основания взято положительное число a, отличное от единицы (если $a = 1$, $1^b = 1$ при любом b).

Логарифмом числа c по основанию a называется такое число b, что $ab = c$, т.е. показатель степени, в которую надо возвести основание, чтобы получить число c:

$$b = \log a\ c.$$

Логарифмы по основанию $a = 10$ используют чаще других, и для них принято сокращенное обозначение: lg.

Равенства $ab = c$ и $b = \log a\ c$ выражают одну и ту же связь между числами a, b и c.

Подставляя в равенство $a^b = c$ запись числа b в виде логарифма, получаем *основное логарифмическое тождество*:

$$a^{\log_a c} = c\,.$$

Подставляя в равенство $b = \log_a c$ выражение в виде степени, получаем еще одно тождество:

$$\log_a a^b = b\,.$$

Свойства степеней и логарифмов тесно связаны между собой. Они фактически выражают одно и то же, только один раз мы обращаем внимание на поведение самих степеней, а другой — на поведение показателей. Запишем *свойства логарифмов, соответствующие свойствам степеней:*

$$a^{b_1}a^{b_2} = a^{b_1+b_2}\,; \qquad \log(c_1c_2) = \log_a c_1 + \log_a c_2\,;$$

$$\frac{a^{b_1}}{a^{b_2}} = a^{b_1-b_2}\,; \qquad \log_a \frac{c_1}{c_2} = \log_a c_1 - \log_a c_2\,;$$

$$(a^b)^k = a^{bk}\,; \qquad \log_a c^k = k\log_a c\,;$$

$$\sqrt[n]{a^b} = a^{b/n}\,; \qquad \log_a \sqrt[n]{c} = \frac{1}{n}\log_a c\,.$$

Свойства степеней отражают свойства умножения чисел. Соответствующие свойства логарифмов выводятся из свойств степеней с помощью основного логарифмического тождества, выражающего определение логарифма.

Выведем первое свойство:

$$\log_a (c_1c_2) = \log_a c_1 + \log_a c_2.$$

Логарифм произведения равен сумме логарифмов.

Обозначим $\log_a c_1 = b_1$, $\log_a c_2 = b_2$. По основному логарифмическому тождеству имеем

$$a^{b_1} = c_1,\ a^{b_2} = c_2.$$

Перемножим эти равенства: $a^{b_1}a^{b_2} = c_1c_2$. По свойству степеней, $a^{b_1}a^{b_2} = a^{b_1+b_2}$, т.е. $c_1c_2 = a^{b_1+b_2}$. По определению логарифма, $b_1 + b_2 = \log_a(c_1c_2)$, тогда

$$\log_a(c_1c_2) = \log_a c_1 + \log_a c_2,$$

что и требовалось доказать.

Первые три свойства показывали поведение степеней и их показателей при умножении, делении и возведении в степень. Извлечение корня n-й степени можно понимать как возведение в степень $1/n$.

С помощью свойств логарифмов можно логарифмировать выражения, составленные с помощью операций умножения, деления и возведения в степень.

Примеры

1. $A = \dfrac{2^5\sqrt{3}}{5^3\sqrt[3]{7^2}}$, $\log_a A = 5\log_a 2 + \dfrac{1}{2}\log_a 3 - 3\log_a 5 - \dfrac{2}{3}\log_a 7$.
2. $A = \dfrac{3a^3\sqrt[3]{b}}{(c+d)^2}$, $\log_a A = \log_a 3 + 3 + \dfrac{1}{3}\log_a b - 2\log_a(c+d)$.
3. $A = 10x^{\log_{10} x}$, $\log_{10} A = 1 + \log_{10} x \cdot \log_{10} x = 1 + (\log_{10} x)^2$.

Иногда приходится искать выражение по его логарифму. Такую операцию называют *потенцированием.*

Примеры

1. $\log_a A = 2\log_a 3 - \dfrac{3}{2}\log_a 2$, $A = \dfrac{3^2}{2^{3/2}} = \dfrac{9}{2\sqrt{2}} = \dfrac{9\sqrt{2}}{4}$.
2. $\log_a A = 1 + 2\log_a b - 3\log_a c$, $A = \dfrac{ab^2}{c^3}$.
3. $\log_a A = \log_a(x-1) + \log_a(x+1) - \dfrac{1}{2}\log_a x$, $A = \dfrac{(x-1)(x+1)}{\sqrt{x}}$.

Запись $t = \log_a b$ имеет смысл лишь при $b > 0$. Поэтому в тождествах, отражающих свойства логарифмов, все выражения, стоящие под знаком логарифма, будем считать положительными. При логарифмировании буквенных выражений надо их раскладывать на множители так, чтобы все множители были положительны. Например, пусть необходимо прологарифмировать выражение $A = x(x-1)$. Сделать это можно лишь тогда, когда $A > 0$, т.е. когда либо $x < 0$, либо $x > 1$. Если $x > 1$, то оба множителя x и $x - 1$ положительны, и мы можем записать

$$\log_a A = \log_a x + \log_a (x-1) \text{ при } x > 1.$$

Если же $x < 0$, то оба множителя отрицательны и A нужно разложить на множители так: $A = (-x)(1-x)$, откуда

$$\log_a A = \log_a (-x) + \log_a (1-x) \text{ при } x < 0.$$

Аналогично, $\log_a x^2 = 2\log_a x$ при $x > 0$ и $\log_a x^2 = 2\log_a(-x)$ при $x < 0$. С помощью модуля это можно записать короче:

$$\log_a x^2 = 2\log_a |x|.$$

В вычислениях в качестве основания a часто берется число $a = 10$. В то же время возникает необходимость проводить вычисления степеней и логарифмов с разными основаниями. Как же связать между собой степени и логарифмы с разными основаниями?

Пусть дана степень (с основанием a) $c = a^b$. Для перехода к новому основанию d запишем число $c = a^b$ в виде d^x при некотором x:

$$c = d^x.$$

Прологарифмируем это равенство по основанию a:

$$\log_a c = x\log_a d,$$

откуда

$$x = \frac{\log_a c}{\log_a d}.$$

Прологарифмировав по основанию d, получим

$$\log_d c = x.$$

Сравнивая эти выражения, находим

$$\log_d c = \frac{\log_a c}{\log_a d}.$$

Это *формула перехода от одного основания к другому*. Таким образом, при изменении основания значения логарифмов изменяются линейно. Коэффициент пропорциональности $\frac{1}{\log_a d}$ часто называют *модулем перехода*.

Отметим простые *следствия формулы перехода:*

1. $\log_a a = \dfrac{1}{\log_a d}$ $(c = a)$.
2. $\log_{a^k} c = \dfrac{\log_a c}{k}$ $(d = a^k)$.
3. $\log_{\frac{1}{2}} x = -\log_a x$ $(k = -1)$.

Контрольные вопросы и задания

1. Обратите внимание на появившиеся в параграфе ключевые слова и обозначения: степень, логарифм, $\log_b a$, модуль перехода.
2. Что называется корнем n-й степени?
3. В чем разница в употреблении слов «корень» и «радикал»?
4. Перечислите основные свойства радикалов.
5. Чему равна нулевая степень числа?
6. Как определяется степень числа с рациональным показателем?
7. Каковы свойства степеней?
8. Что такое логарифм?
9. В чем состоит основное логарифмическое тождество?
10. Перечислите свойства логарифмов.

§ 18. Показательная функция

Порядок роста и убывания функции

Функция — это основной математический инструмент для изучения связей, зависимостей между различными величинами. Чем большим запасом функций мы располагаем, тем шире и богаче наши возможности математического описания окружающего мира.

Вслед за линейными мы подробно изучали квадратичные зависимости. Так, путь при равноускоренном движении квадратично зависит от времени; энергия падающего тела квадратично зависит от его скорости; количество теплоты, выделяемой током, текущим по проводнику, квадратично зависит от силы тока и др.

Степенные зависимости более высокого порядка также встречаются на практике. Например, по закону Стефана—Больцмана, излучательная способность черного тела пропорциональна четвертой степени его температуры. Масса шара является кубической функцией его радиуса.

Если мы изобразим на одном чертеже (рис. 192) графики степенных функций вида $y = x^k$ ($x \geq 0$) при положительных значениях k, то увидим, что чем больше k, тем быстрее при больших значениях x растут эти функции. Степенные функции образуют естественную шкалу, позволяющую сравнивать рост различных функций со стандартными, степенными функциями.

Аналогичная картина наблюдается и с убывающими функциями. Простейшая убывающая функция задается обратно пропорциональной зависимостью $y = \frac{c}{x}$ ($c > 0$, $x > 0$). Изображая на одном чертеже (рис. 193) графики функций $y = \frac{1}{x^k}$ ($x > 0$, $k > 0$), видим, что чем больше k, тем быстрее убывают эти функции при больших значениях x.

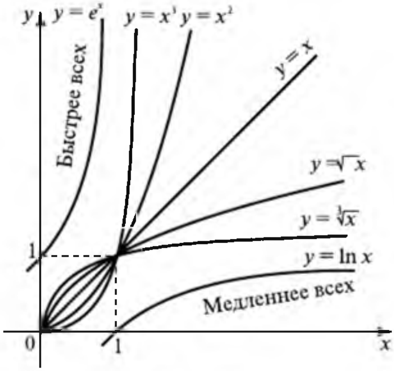


Рис. 192

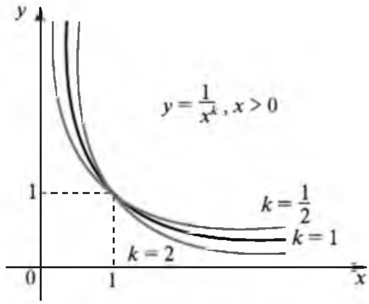


Рис. 193

В естествознании и технике встречаются процессы, рост или затухание которых происходит быстрее, чем у любой степенной функции. С примерами быстро растущих функций человек столкнулся очень давно. В древней легенде об изобретателе шахмат говорится, что он потребовал за первую клетку шахматной доски одно пшеничное зерно, а за каждую следующую — вдвое больше, чем за предыдущую. Человеку трудно представить себе порядок величины $2^{64} - 1$ (общее число зерен, плату за изобретение шахмат). Если грубо заменить $2^{10} = 1024$ на 10^3, то $2^{64} = 2^4 \cdot 2^{60} \approx 16 \cdot 10^{18} = 1{,}6 \cdot 10^{19}$. Достаточно сказать, что расстояние от Земли до Солнца в миллиметрах приблизительно равно $1{,}5 \cdot 10^{14}$, так что, считая диаметр зерна за 1 мм, можно этим зерном 100 тыс. раз уложить путь до Солнца.

Поразительное явление быстрого роста членов геометрической прогрессии, т.е. чисел вида cq^n, отражено во многих старинных задачах. Однако лишь с конца XVII в. стали систематически рассматриваться зависимости типа $y = cq^x$, в которых переменная x принимает не только целые значения. Такие функции называются показательными.

Показательные функции, к изучению которых мы переходим в этой главе, обладают замечательным свойством:

скорость их роста пропорциональна значению самой функции.

Они как костер, который чем больше разгорается, тем больше в него надо подкладывать дров.

Мы знаем, что скорость роста линейной функции постоянна, квадратичной функции линейна и вообще скорость роста (производная) степенной функции, являясь меньшей степенью, растет медленнее, чем сама функция. Необходимость изучения функций, у которых скорость роста (производная) пропорциональна самой функции, возникла с обнаружением различных законов естествознания, таких, как законы размножения, законы радиоактивного излучения, законы движения в тормозящей среде и др. Как эти законы связаны с показательными функциями, мы обсудим в главе, посвященной интегралу.

Исследование показательной функции

Показательной функцией называется функция вида $y = ax$, где a — фиксированное положительное число.

При исследовании показательной функции будем считать, что основание $a \neq 1$, так как при $a = 1$ функция получается постоянной.

Основные свойства показательной функции:

1. Область определения: множество всех вещественных чисел $\mathbb{R}$.
2. Монотонность: при $a > 1$ функция $y = ax$ возрастает, при $0 < a < 1$ функция $y = ax$ убывает.
3. Положительность: значения функции $y = ax$ строго положительны (при любом основании $a > 0$).
4. Область значений: множество всех положительных чисел, т.е. интервал $(0, +\infty)$.

Доказательство свойств 2 и 3. Свойство 1 подчеркивает, что степень a^x определена при любом вещественном x. Доказательство свойств 2 и 3 проводится так: эти свойства проверяются последовательно для натуральных, целых рациональных показателей, а затем уже переносятся на произвольные действительные показатели.

Свойства показательной функции позволяют построить ее график. Графики показательных функций при различных основаниях a приведены на рис. 194.

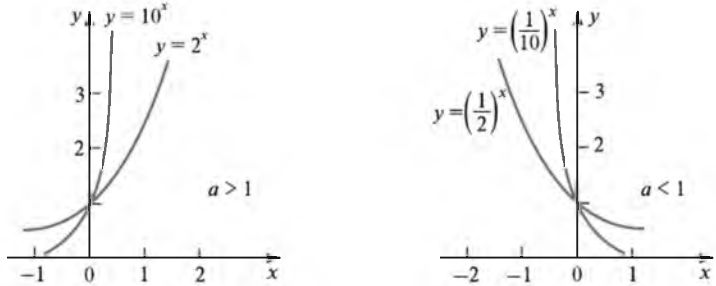


Рис. 194

Пример

Рассмотрим показательную функцию $u = 2^x$. С ростом x значения этой функции возрастают очень быстро. Так $2^{10} \approx 10^3$, $2^{100} \approx 10^{30}$ и т.д. Если $x < 0$, то 2^x быстро приближается к нулю: $2^{-10} \approx 0{,}001$, $2^{-100} \approx 10^{-30}$ и т.д. Это свойство экспоненты — быстро увеличиваться, с одной стороны, и быстро приближаться к нулю — с другой, — хорошо видно на графике.

Вместе с функцией $y = a^x$ показательной функцией считают и функцию вида $y = Ca^x$, где C — постоянная. К такому виду можно привести, например, функцию $y = 2^{x+2}$, сделав преобразование: $2^{x+2} = 2^x \cdot 2^2 = 4 \cdot 2^x$.

Контрольные вопросы и задания

1. Какую функцию называют показательной?
2. Приведите примеры законов естествознания, которые описываются показательными функциями.
3. Перечислите основные свойства показательной функции.
4. Как связана скорость изменения показательной функции (ее производная) с самой функцией?
5. Опишите график показательной функции и ее производной.

§ 19. Логарифмическая функция

Исследование логарифмической функции

Логарифмической функцией называется функция вида $y = \log_a x$, где a — фиксированное положительное число, причем $a \neq 1$.

Основные свойства логарифмической функции

1. Область определения: множество всех положительных чисел, т.е. промежуток $(0, +\infty)$.
2. Монотонность: если $a > 1$, то логарифмическая функция строго возрастает; если $0 < a < 1$, то строго убывает.
3. Область значений: множество всех вещественных чисел $\mathbb{R}$.

Так как определение логарифмов основано на понятии степени, то при доказательстве свойств логарифмической функции используют свойства показательной функции.

Свойство 1 в доказательстве не нуждается, оно опирается на определение логарифма числа x, по которому необходимо, чтобы число x было положительным.

Докажем свойство 2. Для этого сначала рассмотрим случай $a > 1$. Возьмем два положительных числа x_1 и x_2 такие, что $x_1 < x_2$, и докажем, что $\log_a x_1 < \log_a x_2$. Обозначив первое из этих чисел через t_1, второе — через t_2, по определению логарифма получим $a^{t_1} = x_1$, $a^{t_2} = x_2$. Если бы выполнялось неравенство $t_1 \geq t_2$, то по свойству монотонности показательной функции выполнялось бы неравенство $a^{t_1} \geq a^{t_2}$, т.е. $x_1 \geq x_2$. Это противоречит условию. Следовательно, $t_1 < t_2$, что и требовалось доказать. Случай $0 < a < 1$ рассматривается аналогично.

Свойство 3 утверждает, что всякое вещественное число t может быть логарифмом некоторого числа x. Так как степень a^t определена при любом t, то, взяв $x = a^t$, получим $\log_a a^t = t$, что и требовалось доказать.

Графики логарифмических функций при различных основаниях показаны на рис. 195.

Графики функций $y = a^x$ и $y = \log_a x$ симметричны друг другу относительно прямой $y = x$. Действительно, если точка $P(c; d)$ лежит на графике функции $y = a^x$, то $d = a^c$. Но тогда $c = \log_a d$ и точка $Q(d; c)$ лежит на графике функции $y = \log_a x$. Так как точки $P(c; d)$ и $Q(d; c)$ симметричны относительно прямой $y = x$, то симметричны и графики показательной и логарифмической функций.

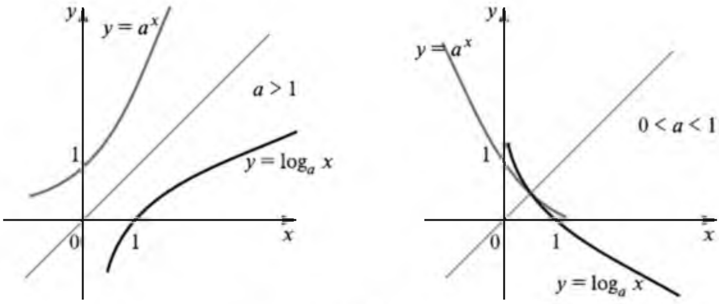


Рис. 195

Вместо логарифмических функций с произвольным основанием удобно рассматривать функции вида $y = c \ln x$. Так как

$$\log_a x = \frac{1}{\ln a} \ln x\,,$$

то указанные функции исчерпывают все логарифмические функции.

Функция $y = \ln x$ растет с ростом x, однако медленнее, чем любая степенная функция вида $y = x^k$ $(k > 0)$, в частности медленнее, чем $y = \sqrt{x}$ (см. рис. 192).

Вычисление логарифмов

Более 300 лет логарифмы использовались для облегчения вычислений. Их основное достоинство — способность сводить умножение к сложению по формуле $\log_a b_1 b_2 = \log_a b_1 + \log_a b_2$. Были составлены обширные таблицы логарифмов чисел, с помощью которых можно легко переходить от чисел к их логарифмам и обратно.

Все таблицы логарифмов до 1950 г. являлись перепечаткой или сокращением таблиц Бриггса. Генри Бриггс (1561—1630) с очень большой точностью (16 знаков после запятой) извлек подряд 57 квадратных корней из 10 и получил значения $10^{1/2}$, $10^{1/4}$, $10^{1/8}$, ..., $10^{1/2^{57}}$. Комбинируя эти значения, он получил густую сетку чисел с известными десятичными логарифмами: $10^{3/4} = 10^{1/2} \cdot 10^{1/4}$ и т.п. После этого десятичный логарифм любого числа x из промежутка $[1;\ 10]$ с хорошей точностью находится округлением до ближайшего известного.

Это огромная работа, и за 300 лет не нашлось никого, кто повторил бы ее. Любопытно, что немного раньше Бриггса таблицу натуральных логарифмов составил Джон Непер (1550—1617).

С появлением компьютера ситуация переменилась. Умножение по-прежнему выполняется дольше, чем сложение, но логарифмирование требует еще больше времени. Поиск числа в таблице — очень дорогая операция для компьютера. Поэтому теперь значение логарифмов как инструмента вычисления резко упало, а с распространением калькуляторов оно сходит на нет. С другой стороны, сами по себе логарифмические зависимости легко обрабатываются и используются при вычислении на компьютере. Например, формула $x^k = \exp(k \ln x)$ служит основным средством возведения в степень (кроме простейших показателей $k = 1, 2, 3$) на всех компьютерах и на калькуляторах.

На современных компьютерах (и на калькуляторах) значения $\ln x$ и e^x вычисляют, пользуясь заранее найденными приближенными формулами. По этим формулам вычисление логарифмов становится довольно простым. Пользователю компьютера никогда не приходится думать о вычислении логарифмов: на всех компьютерах для этого имеются стандартные программы.

Прикладные примеры

1. *Радиоактивный распад.* Изменение массы радиоактивного вещества происходит по формуле $m = m_0 2^{-t/T}$, где m_0 — масса вещества в начальный момент $t = 0$; m — масса вещества в момент времени t; T — некоторая константа, смысл которой мы сейчас выясним.

Вычислим значение m при $t = T$. Так, $m(T) = m_0 2^{-1} = m_0/2$. Это означает, что через время T после начального момента масса радиоактивного вещества уменьшается вдвое. Поэтому число T называют периодом полураспада. Период полураспада радия равен 1600 лет, урана-238 — 4,5 млрд лет, цезия-137 — 31 год, иода-131 — 8 сут.

Закон радиоактивного распада часто записывают в стандартном виде $m = m_0 e^{-t/\tau}$. Связь константы τ с периодом полураспада нетрудно найти:

$$e^{-t/\tau} = 2^{-t/T} \Rightarrow -\frac{t}{\tau} = -\frac{t}{T}\ln 2 \Rightarrow \tau = \frac{T}{\ln 2} = 1{,}45T\,.$$

2. *Рост народонаселения.* Изменение количества людей в стране на небольшом отрезке времени с хорошей точностью описывается формулой $N = N_0 e^{\alpha t}$, где N_0 — количество людей при $t = 0$; N — количество людей в момент времени t; α — некоторая константа.

3. *Барометрическая формула.* Давление воздуха убывает с высотой (при постоянной температуре) по закону $p = p_0 e^{-h/H}$, где p_0 — давление на уровне моря ($h = 0$); p — давление на высоте h; H — некоторая константа, зависящая от температуры. Для температуры 20 °С величина $H \approx 7{,}7$ км.
4. *Формула Циолковского.* Эта формула, связывающая скорость ракеты v с ее массой m, такова: $v = v_1 \ln \frac{m_0}{m}$, где v_1 — скорость вылетающих газов; m_0 — стартовая масса ракеты. Скорость истечения газа при сгорании топлива v_1 невелика (в настоящее время она меньше или равна 2 км/с). Логарифм растет очень медленно, и, для того чтобы достичь космической скорости, необходимо сделать большим отношение $\frac{m_0}{m}$, т.е. почти всю стартовую массу отдать под топливо.
5. *Коэффициент звукоизоляции* стен измеряется по формуле $D = A \lg \frac{p_0}{p}$, где p_0 — давление звука до поглощения; p — давление звука, прошедшего через стену; A — некоторая константа, которая в расчетах принимается равной 20 дБ. Если коэффициент звукоизоляции D равен, например, 20 дБ, то $\lg \frac{p_0}{p} = 1$ и $p_0 = 10p$, т.е. стена снижает давление звука в 10 раз (такую звукоизоляцию имеет деревянная дверь).

Контрольные вопросы и задания

1. Обратите внимание на следующие встречающиеся в тексте ключевые слова и обозначения: логарифм, потенцирование, натуральный логарифм, $\lg x$, $\ln x$, $\log_a x$. Приведите примеры их использования.
2. Каковы основные свойства логарифмов?
3. В чем состоит основное логарифмическое тождество?
4. Какой формулой связаны между собой логарифмы по разным основаниям?
5. Какова область определения логарифмической функции?
6. Что можно сказать о монотонности логарифмической функции?
7. Какова область значений логарифмической функции?
8. Как связаны между собой графики логарифмической и показательной функций?

§ 20. Показательные и логарифмические уравнения и неравенства

Простейшие уравнения

Простейшее показательное уравнение — это уравнение вида $a^x = b$. Пусть основание $a > 0$ и отлично от 1. Так как функция $y = a^x$ строго монотонна, то каждое свое значение она принимает ровно 1 раз. Это означает, что уравнение $a^x = b$ при $b > 0$ имеет одно решение, которое, по определению логарифма, равно $\log_a b$. Если $b \leq 0$, то уравнение $a^x = b$ корней не имеет, так как a^x всегда больше нуля. Если число b записано в виде a^c, т.е. если уравнение представлено в виде $a^x = a^c$, то оно имеет один корень $x = c$. Сформулируем общий результат о решении простейшего показательного уравнения.

Теорема. Пусть $a > 0$ и $a \neq 1$. Уравнение $af(x) = ag(x)$ равносильно уравнению $f(x) = g(x)$.

Доказательство. Докажем, что если $a^{f(x)} = a^{g(x)}$, то $f(x) = g(x)$. Действительно, так как показательная функция строго монотонна, то из равенства ее значений $a^c = a^d$ следует равенство показателей $c = d$. Обратно: если $f(x) = g(x)$, то $a^{f(x)} = a^{g(x)}$.

Примеры

1. $2^x = 1 \Leftrightarrow 2^x = 2^0 \Leftrightarrow x = 0.$
2. $9^x = \dfrac{1}{27} \Leftrightarrow 3^{2x} = 3^{-3} \Leftrightarrow 2x = -3 \Leftrightarrow x = -\dfrac{3}{2}.$
3. $3 \cdot 4^x = 2 \cdot 3^{2x} \Leftrightarrow \dfrac{4^x}{3^{2x}} = \dfrac{2}{3} \Leftrightarrow \left(\dfrac{2}{3}\right)^{2x} = \dfrac{2}{3} \Leftrightarrow 2x = 1 \Leftrightarrow x = \dfrac{1}{2}.$
4. $2 \cdot 5^x = 3^{x+1} \Leftrightarrow \lg 2 + x \lg 5 = (x+1)\lg 3 \Leftrightarrow x(\lg 5 - \lg 3) =$
   $\lg 3 - \lg 2 \Leftrightarrow x = \dfrac{\lg 3 - \lg 2}{\lg 5 - \lg 3}.$
5. $2^{x^2+x-3} = 8^x \Leftrightarrow 2^{x^2-x-3} = 2^{3x} \Leftrightarrow x^2 + x - 3 = 3x \Leftrightarrow x^2 - 2x - 3 = 0 \Leftrightarrow$
   $\Leftrightarrow x_1 = -1 x_2 = 3.$

Простейшее логарифмическое уравнение — это уравнение вида $\log_a x = b$. Оно имеет единственное решение $x = a^b$ при любом b.

Сформулируем общий результат о решении простейшего логарифмического уравнения:

$$\log_a f(x) = \log_a g(x) \Leftrightarrow \begin{cases} f(x) = g(x), \\ f(x) > 0, \\ g(x) > 0. \end{cases}$$

Доказательство. Пусть x — решение уравнения $\log_a f(x) = \log_a g(x)$. Тогда определены логарифмы чисел $f(x)$ и $g(x)$, т.е. эти числа должны быть больше нуля. Потенцируя равенство $\log_a f(x) = \log_a g(x)$, получаем равенство $f(x) = g(x)$.
Обратно: пусть x — решение уравнения $f(x) = g(x)$, причем $g(x) > 0$ и $f(x) > 0$. Тогда равенство $f(x) = g(x)$ можно прологарифмировать и мы получим $\log_a f(x) = \log_a g(x)$, что и требовалось доказать.

Примеры

1. $\log_2 x = 0 \Leftrightarrow x = 2^0 = 1$.
2. $\log_3 x = 1 \Leftrightarrow x = 3$.
3. $\log_{1/2} x = -1 \Leftrightarrow x = \left(\frac{1}{2}\right)^{-1} = 2$.
4. $\log_2 (x - 1) = \log_2 (5 - x) \Leftrightarrow x - 1 = 5 - x;\ x - 1 > 0,\ 5 - x > 0 \Leftrightarrow$ $\Leftrightarrow x = 3$.
   Мы решили уравнение $x - 1 = 5 - x$, а затем проверили, удовлетворяет ли решение условиям $x - 1 > 0$ и $5 - x > 0$. Заметим, что если $f(x) = g(x)$ и $f(x) > 0$, то тогда и $g(x) > 0$, т.е. из двух неравенств достаточно проверить только одно.
5. $\lg (x^2 - x - 3) = \lg x \Leftrightarrow x^2 - x - 3 = x,\ x > 0 \Leftrightarrow x_1 = 3,\ x_2 = -1,\ x > 0 \Leftrightarrow$ $\Leftrightarrow x = 3$.

Простейшие неравенства

Простейшее показательное неравенство — это неравенство вида $a^x > b$ или $a^x < b$ (или $a^x \geq b$, или $a^x \leq b$). Решение такого неравенства нетрудно представить себе графически, построив график показательной функции $y = a^x$ и проведя прямую $y = b$. Рассмотрим для примера два из 16 возможных вариантов.

Пусть $a > 1$ и $b > 0$. Решением неравенства $a^x \geq b$ является промежуток $[\log_a b; +\infty)$, т.е. все числа $x \geq \log_a b$.

Пусть $a > 1$ и $b \leq 0$. Решением неравенства $a^x \geq b$ является множество всех вещественных чисел $\mathbb{R}$.

Примеры

1. $2^x < 1$ *Ответ:* $x < 0$.
2. $\left(\frac{1}{2}\right)^x \geq 4 \Leftrightarrow \left(\frac{1}{2}\right)^x \geq \left(\frac{1}{2}\right)^{-2}$. *Ответ:* $x \leq -2$.
3. $5^x < -5$. *Ответ:* решений нет.
4. $\left(\frac{1}{3}\right)^x > 0$. *Ответ:* x — любое число.
5. $5^x < 3 \cdot 2^x \Leftrightarrow x \lg 5 < \lg 3 + x \lg 2 \Leftrightarrow x (\lg 5 - \lg 2) < \lg 3$. *Ответ*: $x < \frac{\lg 3}{\lg 5 - \lg 2}$ (в решении использовалось то, что $\lg 5 - \lg 2 > 0$).

Можно сказать, что неравенство типа $a^{f(x)} > b$ мы решаем логарифмированием. При логарифмировании неравенств надо помнить два правила: 1) в обеих частях неравенства должны стоять положительные числа; 2) при логарифмировании по основанию $a > 1$ знак неравенства сохраняется, если же $0 < a < 1$, то знак неравенства меняется на противоположный.

Простейшее логарифмическое неравенство — это неравенство вида $\log_a x > b$ (вместо знака $>$ может стоять $<$, $\geq$, $\leq$). Аналогично показательному неравенству здесь также возможно много вариантов. Логарифмическое неравенство решают потенцированием. При этом надо помнить два правила: 1) при переходе от выражения $\log_a f(x)$ к выражению $f(x)$ надо добавлять условие $f(x) > 0$; 2) если $a > 1$, то при потенцировании знак неравенства сохраняется; если же $0 < a < 1$, то знак неравенства меняется на противоположный.

Примеры

1. $\log_2 x > 0 \Leftrightarrow x > 1, x > 0 \Leftrightarrow x > 1$.
2. $\log_2 x < 1 \Leftrightarrow x < 2, x > 0 \Leftrightarrow 0 < x < 2$ или $x \in (0; 2)$.
3. $\log_{1/2}(x-1) \geq -1 \Leftrightarrow x-1 \leq \left(\frac{1}{2}\right)^{-1}$, $x - 1 > 0 \Leftrightarrow 1 < x \leq 3$ или $x \in (1; 3]$.
4. $\log_2 (x^2 - 1) \leq \log_2 (x + 5)$. Сначала учтем условия $x^2 - 1 > 0$ и $x + 5 > 0$. Затем потенцируем: $x^2 - 1 \leq x + 5 \Leftrightarrow x^2 - x - 6 \leq 0 \Leftrightarrow -2 \leq x \leq 3$. Соединяя решения вместе, получим *ответ*: $-2 \leq x < 1$ и $1 < x \leq 3$, или $[-2; -1) \cup (1; 3]$.

Введение новой неизвестной

Основной прием, с помощью которого решают показательные и логарифмические уравнения и неравенства, — это введение новой неизвестной. Поясним этот прием на ряде примеров.

1. *Выражение показательных функций друг через друга.*

Рассмотрим выражения $y_1 = 2^x$, $y_2 = 2^{2x}$, $y_3 = 2^{-x}$, $y_4 = 2^{x/2}$. Все они могут быть алгебраически выражены друг через друга. Например, $y_2 = y_1^2$, $y_3 = \dfrac{1}{y_1}$, $y_4 = \sqrt{y_1}$, $y_2 = \dfrac{1}{y_3^2}$ и т.д. Алгебраическая связь между различными степенями может быть осложнена добавлением в показателе степени постоянных слагаемых: $y_5 = 2^{x+2}$, $y_6 = 2^{2x-1}$, $y_7 = 2^{1-x}$. Однако и сейчас несложно выразить эти выражения, например, через y_1. Получим $y_5 = 2^2 \cdot 2^x = 4y_1$, $y_6 = 2^{-1} \cdot 2^{2x} = \dfrac{1}{2} y_1^2$, $y_7 = 2 \cdot 2^{-x} = \dfrac{2}{y_1}$. К этому полезно напомнить связь между различными основаниями. Например, $\dfrac{1}{2} = 2^{-1}$, $4 = 2^2$, $\sqrt{2} = 2^{1/2}$, $0{,}25 = 2^{-2}$ и т.п. Поэтому выражения $y_8 = \left(\dfrac{1}{2}\right)^x$, $y_9 = 4^{x+1}$, $y_{10} = (\sqrt{2})^{-3x}$, $y_{11} = (0{,}25)^{-x}$ также нетрудно выразить через y_1:

$$y_8 = \frac{1}{y_1},\quad y_9 = 4 \cdot 2^{2x} = 4y_1^2,\quad y_{10} = 2^{-3x/2} = y_1^{-3/2},\quad y_{11} = (2^{-2})^{-x} = 2^{2x} = y_1^2.$$

Если в уравнении или неравенстве встречается несколько показательных функций, то надо все их выразить через одну. Обычно после этого показательное уравнение или неравенство превращается в алгебраическое.

Примеры

1. $3 \cdot 4^x - 2^{x+3} + 8 = 2^{2x+1} - \left(\dfrac{1}{2}\right)^{-1-x}$.

   Пусть $2^x = y$. Тогда $4^x = y^2$, $2^{x+3} = 8y$, $2^{2x+1} = 2y^2$, $\left(\dfrac{1}{2}\right)^{-1-x} = 2y$.

   Уравнение можно записать так: $3y^2 - 8y + 8 = 2y^2 - 2y \Leftrightarrow$ $\Leftrightarrow y^2 - 6y + 8 = 0 \Leftrightarrow y_1 = 4$, $y_2 = 2$. Возвращаясь к неизвестной x, получим $2^x = 4 \Leftrightarrow x = 2$, $2^x = 2 \Leftrightarrow x = 1$. *Ответ*: $x_1 = 2$, $x_2 = 1$.

2. $2^x - \left(\dfrac{1}{2}\right)^{x-1} < 1$.

   Делаем замену $2^x = y$. Неравенство перепишем так: $y - \dfrac{2}{y} < 1 \Leftrightarrow y^2 - y - 2 < 0$ (мы умножили неравенство на y, сохранив знак неравенства, что можно, так как $y = 2^x > 0$) $\Leftrightarrow -1 < y < 2$. Так как $2^x > -1$ верно при всех x, то остается решить неравенство

$2^x < 2 \Leftrightarrow x < 1$. *Ответ*: $x < 1$ (иначе ответ можно записать так: $(-\infty; 1)$).

2. *Выражение логарифмических функций друг через друга.*

Рассмотрим выражения $y_1 = \log_2 x, y_2 = \log_4 x,\ y_3 \log_{1/2} x,\ \ y_4 = \log_{\sqrt{2}} x$. Используя модуль перехода, легко связать эти выражения между собой: $y_2 = \frac{1}{2}y_1,\ y_3 = -y_1, y_4 = 2y_1$.

Свойства логарифмов позволяют по-разному записать связи между выражениями. Например, $y_5 = \log_2 x^2 = 2y_1,\ \ y_6 = \log_2 \sqrt{x} = \frac{1}{2}y_1,\ \ y_7 = \log_2 2x = 1 + y_1,\ \ y_8 = \log_2 \frac{x^3}{2} = -1 + 3y_1$ и т.д.

Если в уравнении или неравенстве встречается несколько логарифмических функций, то надо (если не удается избавиться от логарифмов потенцированием) выразить их через одну и свести логарифмическое уравнение или неравенство к алгебраическому.

Примеры

1. $\frac{1}{5-\lg x} + \frac{2}{1+\lg x} = 1$.

   Делаем замену $\lg x = y$. Получаем уравнение относительно y: $\frac{1}{5-y} + \frac{2}{1+y} = 1 \Leftrightarrow 1 + y + 2\,(5 - y) = (5 - y)(1 + y) \Leftrightarrow y^2 - 5y + 6 = 0 \Leftrightarrow y_1 = 2,\ y_2 = 3$. Возвращаясь к неизвестной x, получаем $\lg x = 2, x = 100$; $\lg x = 3, x = 1000$. *Ответ*: $x_1 = 100, x_2 = 1000$.

2. $\log_3 x \log_9 x \log_{27} x \log_{81} x = \frac{2}{3}$.

   Перейдем к основанию 3. Получим $\log_9 x = \frac{1}{2}\log_3 x$; $\log_{27} x = \frac{1}{3}\log_3 x$; $\log_{81} x = \frac{1}{4}\log_3 x$;. Заменив $\log_3 x$ на y, получим $y \cdot \frac{1}{2} y \times \frac{1}{3} y \cdot \frac{1}{4} y = \frac{2}{3} \Leftrightarrow \frac{1}{24} y^4 = \frac{2}{3} \Leftrightarrow y^4 = 16 \Leftrightarrow y = \pm 2$. Возвращаясь к неизвестной x, получаем $\log_3 x = 2, x = 9$; $\log_3 x = -2,\ x = \frac{1}{9}$. *Ответ*: $x_1 = 9, x_2 = \frac{1}{9}$.

3. $x^{\lg x - 2} \le 1000$.

   Логарифмируя, получаем равносильное неравенство $(\lg x - 2) \times \lg x \le 3$. Положим $\lg x = y$. Получим неравенство $(y - 2)y \le 3 \Leftrightarrow y^2 - 2y - 3 \le 0 \Leftrightarrow -1 \le y \le 3$. Возвращаясь к неизвестной x, получаем $-1 \le \lg x \le 3 \Leftrightarrow \frac{1}{10} \le x \le 1000$. *Ответ*: $\frac{1}{10} \le x \le 1000$, или $\frac{1}{10}$; 1000.

Использование свойства монотонности функций при решении показательных уравнений

Монотонность функций позволяет определить количество корней уравнения, а иногда и найти их значения.

Примеры

1. $2^{2x} = 5 - x$, $x = 1$.
   В левой части уравнения имеем возрастающую функцию, а в правой — убывающую. Следовательно, уравнение не может иметь более одного корня (рис. 196). Один корень можно угадать: $x = 1$. Это число и является окончательным ответом.
2. $4^x - 3^x = 1$.
   Одно решение $x = 1$ легко найти подбором. Докажем, что других корней нет. Перепишем уравнение так: $1 = \left(\frac{3}{4}\right)^x + \left(\frac{1}{4}\right)^x$. В правой части последнего уравнения — сумма убывающих функций. Следовательно, значение $y = 1$ эта сумма может принять только один раз. *Ответ*: $x = 1$.
3. Сколько корней имеет уравнение $e^x = ax$?
   Изобразим схематические графики функций $y = e^x$ и $y = ax$ (рис. 197). При $a < 0$ графики имеют одну точку пересечения.

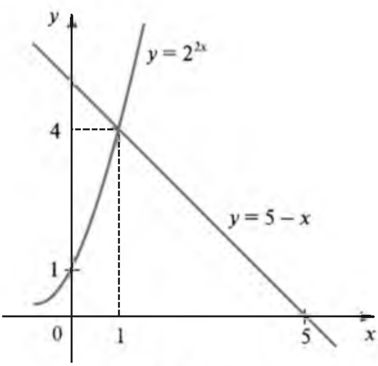


Рис. 196

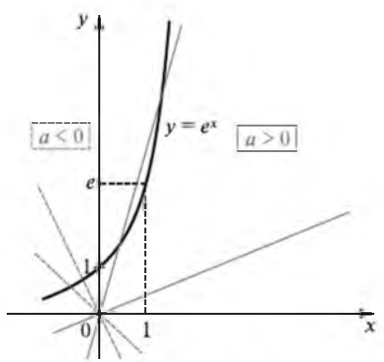


Рис. 197

При $a > 0$ графики могут не пересекаться, касаться друг друга или пересекаться в двух точках. Граничным значением параметра a, при котором происходит разделение основных случаев — две точки пересечения или ни одной, является значение a, при котором прямая $y = ax$ является касательной к графику функции $y = e^x$ в некоторой точке. Это значение параметра равно примерно 2,72 и обозначается буквой e.

Ответ: При $a < 0$ и при $a = e$ — один корень, при $0 \leq a < 0$ нет корней, при $a > e$ два корня.

Контрольные вопросы и задания

1. При каких b показательное уравнение $a^x = b$ имеет корень?
2. Сколько корней имеет уравнение $a^x = b$?
3. Как решать уравнение вида $a^{f(x)} = a^{g(x)}$?
4. Сколько корней имеет уравнение $\log_a x = b$?
5. Как решать уравнение вида $\log_a f(x) = \log_a g(x)$?
6. Какие правила надо соблюдать при логарифмировании неравенств?
7. Какие правила надо соблюдать при потенцировании неравенств?
8. Какова алгебраическая связь между выражениями $y_1 = a^2$ и $y_2 = ak^x$?
9. Какова связь между выражениями: а) $y_1 = \log_a x$ и $y_2 = \log_b x$; б) $y_1 = \log_a x$ и $y_2 = \log_{ak} x$; в) $y_1 = \log_a x$ и $y_2 = \log_{a^{-1}} x$?
10. Вы знаете одно решение уравнения. В каких случаях вы можете доказать, что других решений нет?

ГЛАВА 5

НАЧАЛА СТЕРЕОМЕТРИИ

§ 21. Геометрия Евклида

Введение

Образцом логического совершенства в течение более чем 2000 лет является изложение начал геометрии, предпринятое Евклидом в III в. до н.э. Можно сказать, что это изложение является единственным в истории человечества примером строгой математической теории, с которой полагалось быть знакомым каждому образованному человеку. До сих пор большинство учебников геометрии следует пути, указанному Евклидом, а например, вплоть до недавнего времени английские школьники просто пользовались современным переводом «Начал» Евклида в качестве учебника.

Разумеется, к концу XX в. распространились и иные взгляды. Они относятся к разным вопросам. Вот некоторые из них. Можно ли (и полезно ли) изучать геометрию, отказавшись от ее аксиоматической основы? Нужно ли вообще в рамках общего образования и культуры знакомиться с какой-либо аксиоматической теорией? Если да, то подходит ли для этого геометрия Евклида и нельзя ли найти более простые и доступные примеры? Насколько безупречна сама евклидова геометрия?

Мы не будем касаться этих вопросов. Само построение этой книги положительно отвечает на вопрос, можно ли, изучая математику, обойтись без знакомства с аксиоматическим методом. С другой стороны, мы считаем, что каждый культурный человек должен быть знаком с историей вопроса о евклидовой геометрии, не связывая его, однако, ни с реальным изучением геометрии, ни с целями овладения нового для себя математического метода.

Аксиоматика Евклида

«Начала» Евклида (а точнее, каждая из тринадцати книг, глав, составляющих этот труд) открываются определениями *основных понятий*. Вот первое определение:

Точка есть то, что не имеет частей.

За определениями следуют основные положения, принимаемые без доказательств — *постулаты*, *аксиомы*. (Различие, которое делал Евклид между постулатами и аксиомами, не очень ясно.) Вот несколько примеров.

1. От каждой точки до всякой другой точки можно провести прямую линию.
2. Если прямая, падающая на две прямые, образует по одну сторону внутренние углы, сумма которых меньше двух прямых, то, неограниченно продолженные, эти прямые пересекаются по эту сторону (рис. 198).

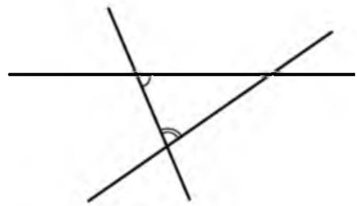


Рис. 198

Это и есть знаменитый пятый постулат, равносильный аксиоме о единственности параллельной.

Затем с помощью основных понятий и аксиом чисто логическим путем доказываются теоремы (предложения). Так, в качестве уже четвертого по счету предложения Евклид доказывает «первый признак равенства треугольников».

Разумеется, Евклид пользуется многим из того, чего нет на самом деле в аксиомах (например, там нет ничего о наложении фигур, которое часто используется как своего рода мысленный эксперимент). Однако за исключением нескольких частностей редакционного и языкового характера, уровень строгости Евклида считался вполне удовлетворительным вплоть до конца XIX в.

Современная аксиоматика евклидовой геометрии

Как уже было сказано, почти все школьные учебники геометрии воспроизводят ту или иную аксиоматику, причем, как правило, в начале курса. При этом стараются сделать список возможно более простым и в то же время удобным для доказательства теорем. Образцом для такого построения, учитывающим достижения и язык математики, сложившиеся к концу XIX в., служит замечательная (хотя и не очень простая) система аксиом немецкого математика Гильберта, созданная им в 1899 г.

Гильберт выделяет три системы неопределяемых (основных) объектов: точки, прямые и плоскости. Затем постулируются «отношения» между ними (принадлежность, находиться между, быть равными, конгруэнтными). Эти положения образуют три группы аксиом. Например, во вторую группу аксиом («аксиомы порядка») попало следующее утверждение, на которое обратил внимание немецкий геометр Паш в 1882 г. как на необходимую аксиому: «Если прямая входит внутрь треугольника через одну из его сторон, но не через его вершину, то она должна выходить из него через другую сторону».

Четвертая группа состоит из одной аксиомы — аксиомы о параллельности: «Через всякую точку, лежащую вне прямой a, проходит не более одной прямой, параллельной a».

В пятую группу включены две аксиомы о непрерывности, в том числе так называемая аксиома Архимеда: «При любых отрезках a и b можно отложить вдоль b отрезки, равные a, столько раз, что они покроют отрезок b».

Неевклидова геометрия

В течение 2000 лет никто не сомневался (и не сомневается по сей день) в ценности аксиоматики Евклида. Единственный вопрос, к которому постоянно возвращались люди (как профессиональные математики, так и бесчисленные любители), состоял в следующем: нельзя ли доказать, вывести пятый постулат Евклида из остальных аксиом как некоторую теорему. По этому вопросу были написаны тысячи книг и статей, но лучшее, что удавалось сделать, это заменить аксиому о параллельности другим утверждением, которое казалось гораздо более очевидным (и потому часто упускалось из виду), но которое оказывалось совершенно равносильным пятому постулату.

К середине XIX в. стало ясно, что пятый постулат независим от остальной аксиоматики Евклида в том замечательном смысле, что, добавив к этой остальной системе аксиому, отрицающую пятый постулат (например, в такой форме, что через всякую точку проходит по крайней мере две прямых, параллельных данной), мы получим новую, *непротиворечивую* систему, в которой можно получить столь же далекие и содержательные теоремы, как те, которые были получены в геометрии Евклида.

Такая система, в которой выполняется указанный список аксиом (включая отрицание пятого постулата), стала называться «неевклидовой геометрией». Впервые такую систему четко описал замечательный русский математик Н.И. Лобачевский (1792—1856), который выступил с докладом в Казанском Университете в 1826 г., а через четыре года подробно изложил свою теорию. Независимо от Лобачевского в 1832 г. венгерский математик Я. Бойяи опубликовал работу с аналогичным содержанием. Еще через 40 лет были построены примеры поверхностей, на которых выполняется геометрия Лобачевского.

От геометрии к логике

Смысл движения от Евклида к Лобачевскому и далее к Гильберту (или, по крайней мере, один из этих смыслов) состоит в освобождении от геометрической наглядности. В системе Гильберта не нужно знать, как «выглядят» точки и прямые. С ними можно (и нужно) обращаться как с объектами, о которых известно лишь то, что описано в аксиомах. Тем самым, исходя из аксиом, мы получаем новые результаты лишь с помощью логики. Такой взгляд порождает два новых вопроса. Прежде всего о самой геометрии. Получение в качестве следствий аксиом большого числа достаточно глубоких утверждений (как это, например, было сделано Лобачевским) само по себе не доказывает непротиворечивости построенной системы. Уже к концу XIX в. стало ясно, что доказывать непротиворечивость новых систем можно с помощью *моделей*, реализующих аксиомы системы. Так, Гильберт указывает модель построения евклидовой геометрии с помощью чисел. Другой вопрос связан с анализом самой логики, что и было предпринято с большой интенсивностью в XX в.

§ 22. Взаимное расположение прямых и плоскостей

Прямые в пространстве

Точка, прямая, плоскость — с этими словами у нас связаны определенные представления. Мы говорим: «прямой, как натянутая струна», «пересечь улицу по прямой», «луч света движется по прямой». Линии пересечения стен и потолков, траектории свободно падающих тел служат наглядными примерами прямых линий.

Со словом «плоскость» у нас ассоциируется поверхность стола, стены, пола. Плоской в нашем представлении является поверхность озера в безветренный день или гладь хорошо залитого катка.

Дотронемся мелом до доски. Остается след — точка. Посмотрим на ночное небо — звезда нам кажется светящейся точкой. Взглянем на угол комнаты — это тоже точка.

Нашей целью является приведение в систему наших пространственных, стереометрических представлений.

Основные понятия стереометрии: точка, прямая, плоскость.

Аксиомы стереометрии

1. Каждая прямая содержит, по меньшей мере, две точки.
2. Через любые две точки проходит одна и только одна прямая.
3. Каждая плоскость содержит, по меньшей мере, три точки.
4. Через любые три точки, не лежащие на одной прямой, проходит одна и только одна плоскость.
5. Если две точки прямой принадлежат некоторой плоскости, то вся прямая содержится в этой плоскости.
6. Если две плоскости имеют общую точку, то множество всех их общих точек есть прямая (т.е. плоскости пересекаются по прямой).
7. Через любую точку, не лежащую на данной прямой, проходит одна и только одна прямая, параллельная данной.

Исходным объектом стереометрии, ее вселенной, универсумом является *пространство*. Мы будем считать, что пространство состоит из точек. При этом в пространстве выделены множества двух типов — прямые и плоскости, удовлетворяющие перечисленным аксиомам.

Прежде всего, уточним представления о взаимном расположении прямых и плоскостей в пространстве.

Две прямые, которые не лежат в одной плоскости, называются скрещивающимися.

Две прямые называются параллельными, если они лежат в одной плоскости и не имеют общих точек (не пересекаются).

Две прямые, имеющие ровно одну точку, называются пересекающимися (рис. 199).

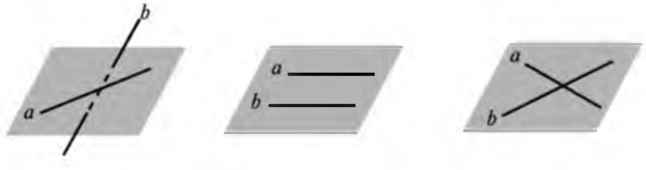


Рис. 199

Для пары прямых в пространстве имеются две возможности: либо через них можно провести плоскость, либо нельзя. Если плоскость провести нельзя, то прямые — скрещивающиеся. Рассмотрим первую возможность.

Пусть прямые лежат в одной плоскости. Тогда они либо имеют общую точку (пересекаются), либо нет (параллельны).

Обратим внимание на логику рассуждений, которые мы применяли при разборе вариантов расположения двух прямых в пространстве. Сначала мы разбили пары прямых на типы по такому признаку: прямые лежат или не лежат в одной плоскости. Ясно, что это два взаимно исключающие друг друга случая, исчерпывающие все мыслимые возможности. Затем мы рассмотрели пары прямых, лежащих в одной плоскости, и прибегли к дополнительному признаку: прямые не имеют общих точек или прямые имеют одну общую точку.

Сколько общих точек могут иметь две прямые? Из планиметрии нам известно, что через две (различные) точки можно провести только одну прямую, поэтому две (различные) прямые не могут иметь более одной общей точки. Это доказывает, что две различные прямые могут быть либо скрещивающимися, либо параллельными, либо пересекающимися.

Примеры

1. На рисунке 200 изображен разрез обычной комнаты. Плоскости стен, пола и потолка пересекаются по прямым. Три из этих прямых выделены и обозначены a, b и c. Прямые a и b пересекаются, прямые a и c параллельны, прямые b и c скрещиваются. Укажите на этом рисунке другие пары параллельных, пересекающихся и скрещивающихся прямых.
2. На рисунке 201 изображен токарный резец. Его кромки — отрезки прямых линий. Разберите самостоятельно расположение этих прямых.
3. Выберем четыре произвольные точки A, B, C и D, не лежащие в одной плоскости (рис. 202). Проведем прямые AB и CD. Они

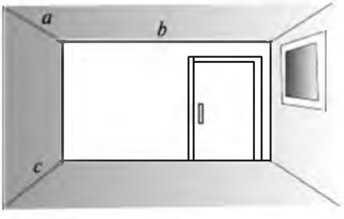


Рис. 200

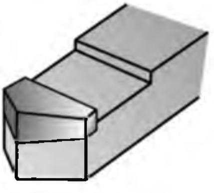


Рис. 201

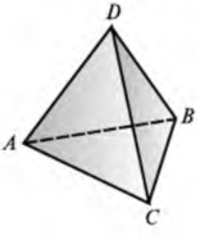


Рис. 202

скрещиваются, так как если бы они лежали в одной плоскости, то все данные точки лежали бы в одной плоскости, что противоречит условию. Точки A, B, C и D определяют три пары скрещивающихся прямых AB и CD, AC и BD, AD и BC. На этих прямых лежат ребра треугольной пирамиды $ABCD$.

Каких пар прямых больше — параллельных, пересекающихся или скрещивающихся? Вопрос поставлен неточно — разумеется, есть бесконечно много пар и параллельных, и пересекающихся, и скрещивающихся прямых. Однако он все же имеет смысл. Две наугад взятые прямые скорее всего не будут лежать в одной плоскости, а значит, будут скрещиваться. Пересечение или параллельность двух прямых в пространстве — это в каком-то смысле специальный, частный случай их расположения, а скрещивание — общий. Про скрещивающиеся прямые говорят иногда, что они находятся в общем положении.

Признак скрещивающихся прямых

Теорема 1. Пусть даны две прямые a и b. Если можно найти такую плоскость α, которая содержит прямую a и пересекает прямую b в точке, не лежащей на прямой a, то прямые a и b скрещиваются (рис. 203).

Доказательство. Будем доказывать теорему от противного. Прямые скрещиваются — это означает, что они не лежат в одной плоскости. Противоположное утверждение означает, что прямые a и b лежат в одной плоскости. Эта плоскость должна совпадать с плоскостью α. Действительно, в ней лежат прямая a и вся прямая b, а значит, та точка A прямой b, о которой говорилось в условии теоремы. Однако ясно, что прямая a и не лежащая на ней точка A однозначно определяют плоскость. Значит, вся прямая b должна лежать в плоскости α, что противоречит условию. Следовательно, исходное предположение о том, что прямые a и b лежат в одной плоскости, неверно. Теорема доказана.

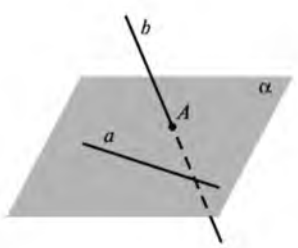


Рис. 203

С помощью условных знаков теорему 1 можно записать так:

Объекты	*Взаимное расположение*
прямые a, b	$a \in \alpha$;
плоскость α	$A \in b$, $A = b \cap \alpha$;
точка A	$A \notin a$.

Доказать: $a \mathbin{\dot{-}} b$

Признак скрещивающихся прямых является первой доказанной нами теоремой стереометрии. Откуда получается первая теорема, на что опирается ее доказательство, какие аргументы мы привели в защиту доказываемого утверждения? Во-первых, нам помогает логика. Мы понимаем, что две прямые либо лежат в одной плоскости, либо нет. Если мы привели к противоречию предположение о том, что они лежат в одной плоскости, значит, они не лежат в одной плоскости, т.е. скрещиваются (по определению скрещивающихся прямых). Однако одной логики нам было мало. В доказательстве использованы некоторые дополнительные свойства прямых и плоскостей, которые мы никак не аргументировали, так как они хорошо соответствуют нашим представлениям о расположении прямых и плоскостей в пространстве. В рассуждении использовано, что прямая и не лежащая на ней точка однозначно задают плоскость.

Сформулируем более полно свойство пространства, использованное в теореме 1, которое можно вывести из аксиом стереометрии.

Свойство 1

Через прямую и не лежащую на ней точку можно провести плоскость, и притом только одну.

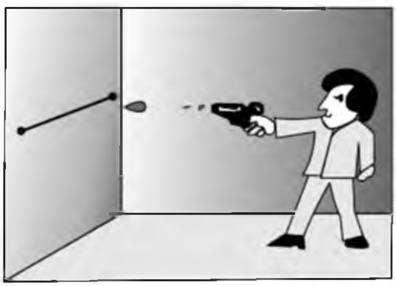


Рис. 204

Пример

Представим себе, что на стене проведена прямая (рис. 204), а мы стреляем в стену (считаем, что пуля полетит по прямой). При попадании в прямую траектория пули будет пересекать ее, а при промахе — скрещиваться с ней (применение признака требует, чтобы пуля при промахе все-таки попала в стену). Ясно, что промахнуться неизмеримо легче, чем попасть в тонкую линию. Это еще раз подчеркивает, насколько скрещивающихся прямых больше, чем пересекающихся.

Прямая и плоскость

Прямая и плоскость называются пересекающимися, если у них только одна общая точка.

Прямая и плоскость называются параллельными, если они не имеют общих точек (рис. 205).

Начнем классификацию пар, состоящих из плоскости и прямой, лежащей вне этой плоскости, со следующего признака: прямая и плоскость имеют общие точки или прямая и плоскость не имеют общих точек. Во втором случае прямую и плоскость мы назвали параллельными. Вернемся к первому случаю. Сколько может быть общих точек у прямой и плоскости? Наши наглядные представления подсказывают, что если у прямой и плоскости более одной общей точки, то вся прямая целиком лежит в данной плоскости (если ровную линейку в двух точках приложить к хорошо отшлифованной плоской поверхности, то линейка целиком ляжет на эту поверхность). Таким образом, наша классификация взаимных расположений прямой и плоскости опирается на аксиому 5:

Если две различные точки прямой лежат в некоторой плоскости, то вся прямая лежит в этой плоскости (рис. 206).

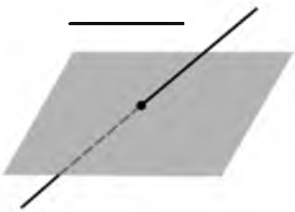


Рис. 205

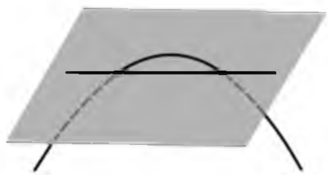
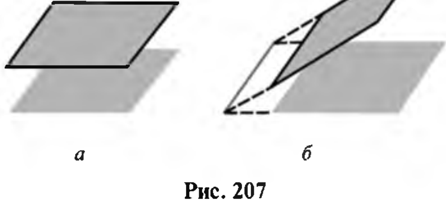


Рис. 206

Плоскости в пространстве

Две плоскости называются параллельными, если они не имеют общих точек (рис. 207, *а*).
Две плоскости называются пересекающимися, если множество их общих точек есть прямая (рис. 207, *б*).



а *б*

Рис. 207

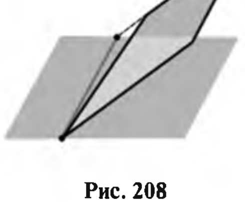


Рис. 208

Снова рассмотрим возможные варианты. Две плоскости либо не имеют общих точек (параллельны), либо имеют общие точки. Наглядно ясно, что если две различные плоскости имеют общие точки, то общие точки заполняют целую прямую. Это и составляет содержание аксиомы 6:

Если две различные плоскости имеют хотя бы одну общую точку, то они пересекаются по прямой, т.е. множество их общих точек есть прямая (рис. 208).

На рисунке 209 изображена часть комнаты. Плоскость пола обозначена через α. Из выделенных трех прямых прямая a лежит в плоскости α, прямая b пересекается с плоскостью α, прямая c параллельна плоскости α. Как расположены по отношению к плоскости α другие прямые, изображенные на рис. 209?

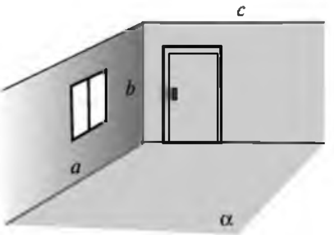


Рис. 209

Возьмем наугад плоскость и прямую в пространстве. Какое их взаимное расположение будет более вероятным? Скорее всего прямая и плоскость будут пересекаться. Про пересекающиеся плоскости и прямую говорят, что они находятся в общем положении. Другие случаи взаимного расположения плоскости и прямой в пространстве являются более частными, специальными.

Плоскости пола и потолка параллельны друг другу. Плоскости стен пересекают плоскость пола. Как расположены между собой плоскости стен?

Контрольные вопросы и задания

1. Обратите внимание на следующие ключевые слова, появившиеся в тексте: точка, прямая, плоскость, скрещивающиеся, параллельные, пересекающиеся. Проведите самостоятельно классификацию расположения пары прямых в пространстве, начиная с признака: прямые имеют общие точки или прямые не имеют общих точек.
2. Дайте определения:
   а) скрещивающихся прямых;
   б) пересекающихся прямых;
   в) пересекающихся прямой и плоскости;
   г) параллельных прямой и плоскости;

д) параллельных плоскостей;
е) пересекающихся плоскостей.

3. Сформулируйте признак скрещивающихся прямых.
4. Какие свойства пространства были приняты без доказательства при анализе взаимного расположения прямых и плоскостей?
5. Какое взаимное расположение вы считаете общим, наиболее часто встречающимся:
   а) для двух прямых;
   б) для прямой и плоскости;
   в) для двух плоскостей?

§ 23. Признаки параллельности

Признак параллельности прямой и плоскости

Теорема 2. Если прямая, не лежащая в плоскости, параллельна прямой на плоскости, то исходная прямая параллельна плоскости.

Более подробно эта теорема может быть сформулирована следующим образом.

Пусть прямая a не лежит в плоскости α. Если в плоскости α есть прямая, параллельная прямой a, то прямая a параллельна плоскости α (рис. 210).

Остановимся на логической структуре доказательства теоремы 2. В условии теоремы указано расположение прямых a и b относительно плоскости α ($a \notin \alpha$, $b \in \alpha$). Сама теорема представляет собой условное утверждение: если $a \parallel b$, то $a \parallel \alpha$. При доказательстве строится отрицание заключения: пусть неверно, что $a \parallel \alpha$. С учетом условия $a \notin \alpha$ это означает, что прямая a и плоскость α пересекаются. Это утверждение приводится к противоречию с условием $a \parallel b$. Такой тип доказательства называют «доказательством от противного».

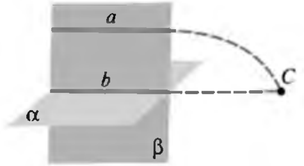


Рис. 210

Запишем кратко условие теоремы.

Объекты	*Взаимное расположение*
прямая a	$a \notin \alpha$;
плоскость α	$b \in \alpha$;
прямая b	$a \parallel b$.

Доказать: $a \parallel \alpha$

Доказательство. Каковы логически возможные случаи расположения прямой a и плоскости α? Мы знаем, что либо a лежит в плоскости α, либо a пересекает плоскость α, либо a параллельна α. Первая возможность отвергнута условием. Достаточно опровергнуть вторую возможность.

Проведем через параллельные прямые a и b плоскость. Обозначим ее β. Если прямая a пересекает плоскость α в некоторой точке C, то эта точка — общая для плоскостей β и α. Но все общие точки различных плоскостей α и β лежат на прямой b. Следовательно, точка C лежит на прямой b, но это противоречит параллельности прямых a и b. Теорема доказана.

Признак параллельности двух прямых

Теорема 3. Если плоскость проходит через прямую, параллельную другой плоскости, и пересекает эту плоскость, то линия пересечения параллельна исходной прямой.

Введем обозначения и более подробно сформулируем теорему. Нам даны некоторая прямая a и параллельная ей плоскость β. Другая плоскость α проходит через прямую a и пересекает плоскость β по некоторой прямой b. Теорема утверждает, что прямые a и b параллельны.

Запишем кратко условие теоремы.

Объекты	*Взаимное расположение*
прямые a, b	$a \in \alpha$;
плоскости α, β	$\alpha \parallel \beta$.
Доказать: $a \parallel b$	$b = \alpha \cap \beta$.

Доказательство. По условию, прямые a и b (рис. 211) лежат в одной плоскости α. Следовательно, они либо параллельны, либо пересекаются (вспомним классификацию пар прямых). Если бы они пересекались, то их общая точка C лежала бы одновременно на прямой a и в плоскости β (так как прямая b лежит в плоскости β). Это противоречит тому, что прямая a параллельна плоскости β. Теорема доказана.

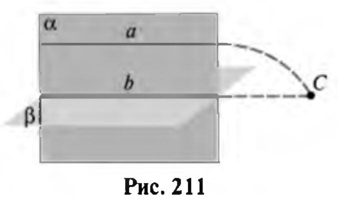


Рис. 211

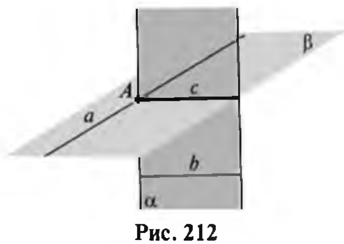


Рис. 212

Задача

Даны две скрещивающиеся прямые. Провести через одну из них плоскость, параллельную другой (рис. 212).

Обозначим данные скрещивающиеся прямые через a и b. Возьмем на прямой a произвольную точку A и проведем плоскость α через прямую b и точку A.

Проведем в плоскости α через точку A прямую c, параллельную прямой b. Через эти две пересекающиеся прямые a и c проведем плоскость β. Эта плоскость будет параллельна прямой b. Почему? Как вы думаете, сколько есть плоскостей, проходящих через прямую a и параллельных прямой b?

Признак параллельности двух плоскостей

Теорема 4. Если каждая из двух пересекающихся прямых одной плоскости параллельна другой плоскости, то эти плоскости параллельны (рис. 213).

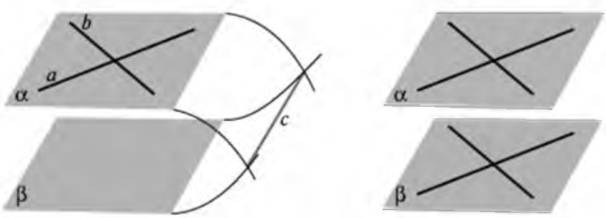


Рис. 213

Запишем кратко условие теоремы.

Объекты	*Взаимное расположение*
плоскости α, β	$a \in \alpha$, $b \in \alpha$;
прямые a, b	$a \times b$, $a \parallel \beta$, $b \parallel \beta$.

Доказать: $\alpha \parallel \beta$.

Доказательство. Будем рассуждать от противного. Пусть плоскости α и β не параллельны. Тогда они пересекаются по некоторой прямой, обозначим ее c. Все три прямые a, b и c лежат в плоскости α.

Так как прямые a и b параллельны плоскости β, то, по теореме 2, прямая c, как линия пересечения плоскостей α и β, параллельна прямым a и b.

Итак, через точку пересечения прямых a и b в плоскости α проходят две прямые (a и b), параллельные прямой c. Это противоречит аксиоме о параллельных из планиметрии. Полученное противоречие доказывает теорему.

При решении задач часто используют другую формулировку теоремы 4.

Если две пересекающиеся прямые одной плоскости соответственно параллельны двум прямым другой плоскости, то эти плоскости параллельны (см. рис. 213).

Иными словами, если в двух плоскостях можно найти по углу с соответственно параллельными сторонами, то эти плоскости параллельны. Выведите самостоятельно сформулированное утверждение из теоремы 4, используя признак параллельности прямой и плоскости.

Параллельность трех прямых

Теорема 5. Если две прямые параллельны третьей, то они параллельны между собой (рис. 214).

Запишем кратко условие теоремы.

Объекты	*Взаимное расположение*
прямые a, b, c	$a \parallel c$ и $b \parallel a$.

Доказать: $b \parallel c$.

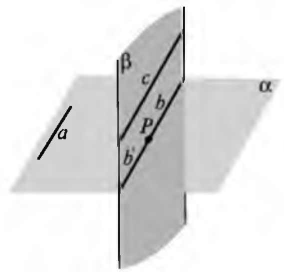


Рис. 214

Доказательство. Если все прямые a, b и c лежат в одной плоскости, то получается известная в планиметрии теорема.

Пусть прямые a, b и c не лежат в одной плоскости. Нам нужно доказать, что прямые c и b параллельны, т.е. лежат в одной плоскости и не имеют общих точек.

Как доказать, что c и b лежат в одной плоскости? Здесь никакое применение доказанных ранее теорем не поможет, так как ни одна из них не содержит вывода, что прямые по каким-то причинам лежат в одной плоскости. Однако плоскость можно провести через прямую и не лежащую на ней точку (свойство 1). Поэтому поступим так. Возьмем на прямой b произвольную точку P и проведем плоскость β через нее и прямую c. Остается доказать, что пересечение плоскости β с плоскостью α, проведенной через прямые b и a, есть прямая b.

Прямая a не лежит в плоскости β и параллельна прямой c, лежащей в плоскости β. Согласно теореме 3, прямая a параллельна плоскости β. Из этой же теоремы следует, что линия пересечения b' плоскостей α и β параллельна прямой a. Поскольку же b, b', a лежат в одной плоскости и прямая b, имеющая с b' общую точку, тоже параллельна a, получаем $b = b'$. Теорема доказана.

Типовые задачи на построение и их разрешимость

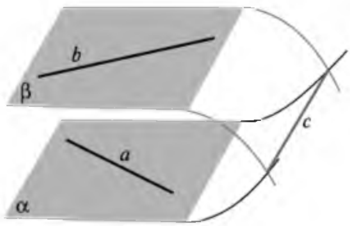


Рис. 215

Задача 1

Даны две скрещивающиеся прямые. Провести через каждую из них по плоскости так, чтобы эти плоскости получились параллельными (рис. 215).

Мы уже умеем проводить через одну из скрещивающихся прямых плоскость, параллельную другой.

Пусть даны две скрещивающиеся прямые a и b. Проведем через прямую a плоскость α, параллельную b, а через b — плоскость β, параллельную a.

Докажем, что плоскости α и β параллельны. Если бы они пересекались, то линия их пересечения — прямая c — была бы параллельна каждой из прямых a и b (дайте обоснование). По теореме 5, прямые a и b должны были бы быть параллельными, но известно, что они скрещиваются. Задача решена.

Построение сечений многогранников

Признаки параллельности применяются для построения сечений кубов, призм, пирамид и других многогранников. Для задания секущей плоскости можно использовать следующие элементы:

1) три точки, не лежащие на одной прямой;
2) прямая и не лежащая на ней точка;
3) пара пересекающихся прямых;
4) пара параллельных прямых.

Каждый из этих случаев определяет ровно одну плоскость.

Плоскость сечения пересекает грани многогранника по отрезкам прямых. Обычно для изображения сечения достаточно построить все эти отрезки. Иногда полезно построить пересечение с диагональными плоскостями (в призме) или другими выделенными плоскостями.

Задача 2

Построить сечение куба плоскостью, проходящей через середины двух смежных сторон основания и через центр куба.

Дано: куб $ABCDA'B'C'D'$, K — середина AB, L — середина AD, O — центр куба. Указанные три точки однозначно определяют плоскость сечения.

Соединим точки K и L отрезком. Проведем диагональное сечение $B'D'DB$ (рис. 216).

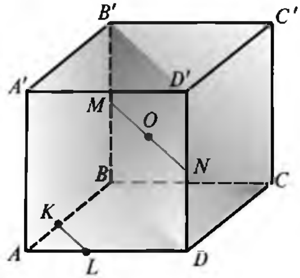


Рис. 216

Центр куба O является центром прямоугольника $B'D'DB$. Плоскость сечения проходит через KL и пересекает плоскость диагонального сечения. KL параллельна плоскости диагонального сечения, так как $KL \parallel BD$ (как средняя линия треугольника ABD) — признак параллельности прямой и плоскости. Линия пересечения должна быть параллельна прямой KL (признак параллельности двух прямых) и, следовательно, параллельна BD (по теореме о параллельности трех прямых). Проводим через точку O прямую $MN \parallel BD$. Мы получим точки сечения, расположенные на ребрах BB' и DD'.

Рассмотрим второе диагональное сечение $AA'C'C$ (рис. 217). Оно пересекает отрезок KL в точке P (его середина) и проходит через центр O. Плоскость сечения пересекает эту диагональную плоскость по прямой PO. Пусть Q — точка пересечения этой прямой с верхним основанием (с отрезком $A'C'$). Рассмотрим, как пересекает искомая плоскость сечения плоскость верхнего основания. Так как прямая $KL \parallel BD$, $BD \parallel B'D'$, то KL параллельна плоскости этого основания и линия пересечения будет тоже параллельна всем этим прямым. Проводя через точку Q прямую $RS \parallel B'D'$, получим недостающие точки сечения (рис. 218).

В сечении получился шестиугольник. Вторую половину построения можно было сократить, заметив, что сечение должно быть симметрично относительно центра. Равны ли между собой стороны построенного шестиугольника? Будет ли этот шестиугольник правильным?

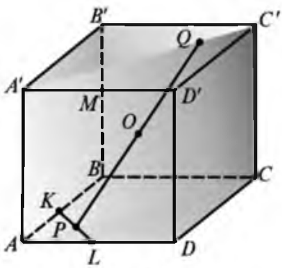


Рис. 217

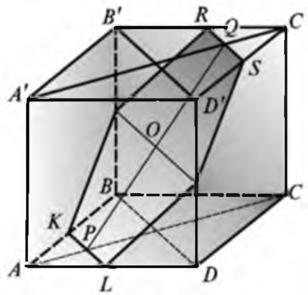


Рис. 218

Плоскость сечения может быть задана условиями параллельности некоторым прямым или плоскостям.

1. Даны прямая и плоскость. Ставится задача провести плоскость через данную прямую параллельно данной плоскости. Эта задача разрешима только в том случае, когда данная прямая и плоскость параллельны. В этом случае решение единственно.
2. Даны плоскость и точка вне ее. Всегда можно провести плоскость через данную точку параллельно данной плоскости. Решение единственно.
3. Даны две скрещивающиеся прямые. Провести плоскость через одну из них параллельно другой (см. задачу 1).
4. Даны две прямые и точка. Требуется построить плоскость, проходящую через заданную точку, параллельную обеим прямым. Исследуйте самостоятельно разрешимость этой задачи.

Задача 3

В четырехугольной пирамиде $SABCD$ провести сечение через точки P и Q, лежащие на боковых ребрах SA и SC, параллельно диагонали основания BD.

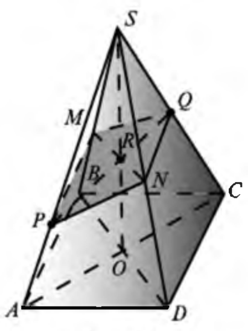


Рис. 219

Проведем диагональные сечения SAC и SBD. Они пересекаются по прямой SO, где O — точка пересечения диагоналей основания (рис. 219).

Соединим точки P и Q. Отрезок PQ пересекает SO в некоторой точке R. Плоскость сечения, по условию, должна быть параллельна прямой BD. Поэтому она пересекает плоскость SBD по прямой, параллельной BD (признак параллельности двух прямых). Проведем через точку R в плоскости SBD прямую, параллельную BD. Получим недостающие точки M и N сечения. Задача решена.

Контрольные вопросы и задания

1. Сформулируйте признак параллельности прямой и плоскости.
2. С помощью каких теорем можно доказать параллельность двух прямых?
3. В чем состоит признак параллельности двух плоскостей?
4. Если две ... параллельны третьей, то они параллельны между собой. Что можно подставить вместо многоточия?
5. Какие элементы можно использовать для задания плоскости?

§ 24. Применение векторов

Параллельность и перпендикулярность прямых и плоскостей

Использование векторов упрощает проверку параллельности и перпендикулярности прямых и плоскостей. Мы уже отмечали в главе 2, что угол между двумя прямыми можно определить через угол между направляющими векторами этих прямых. Если $\vec{v}_1$ и $\vec{v}_2$ — направляющие векторы двух прямых l_1 и l_2, то

$l_1 \parallel l_2 \Leftrightarrow \vec{v}_1$ коллинеарен $\vec{v}_2 \Leftrightarrow$ существует число λ такое, что $\vec{v}_1 = \lambda\, \vec{v}_2$;

$l_1 \perp l_2 \Leftrightarrow \vec{v}_1 \perp \vec{v}_2 \Leftrightarrow \vec{v}_1 \cdot \vec{v}_2 = 0.$

Признак перпендикулярности прямой и плоскости

В определении прямой, перпендикулярной плоскости, содержится требование, чтобы эта прямая была перпендикулярна каждой прямой данной плоскости. Прямых в плоскости бесконечно много, и поэтому непосредственно пользоваться этим определением нельзя. Оказывается, что если данная прямая перпендикулярна всего двум пересекающимся прямым в плоскости, то она уже перпендикулярна любой другой прямой в этой плоскости.

Теорема. Если прямая перпендикулярна двум пересекающимся прямым некоторой плоскости, то она перпендикулярна этой плоскости.

Запишем кратко условие теоремы.

Объекты	*Взаимное расположение*
плоскость α	$l_1 \in \alpha$, $l_2 \in \alpha$;
прямые l, l_1, l_2	$l_1 \times l_2$;
Доказать: $l \perp \alpha$	$l \perp l_1$, $l \perp l_2$.

Доказательство. По определению, прямая l перпендикулярна плоскости α, если она перпендикулярна любой прямой этой плоскости. Выберем произвольную прямую $m \in \alpha$ и покажем, что $l \perp m$. Для каждой прямой выберем направляющий вектор.

Прямая	*Направляющий вектор*
l	$\mathbf{v}$;
l_1	$\mathbf{v}_1$;
l_2	$\mathbf{v}_2$;
m	$\mathbf{u}$.

Так как прямые l_1 и l_2 пересекаются (рис. 220), то векторы $\mathbf{v}_1$ и $\mathbf{v}_2$ не коллинеарны и поэтому вектор $\mathbf{u}$, лежащий в плоскости α, можно разложить по этим векторам: $\mathbf{u} = \lambda_1\mathbf{v}_1 + \lambda_2\mathbf{v}_2$.

Для проверки того, что $l \perp m$, вычислим скалярное произведение векторов $\mathbf{v}$ и $\mathbf{u}$:

$$\mathbf{vu} = \mathbf{v}(\lambda_1\mathbf{v}_1 + \lambda_2\mathbf{v}_2) = \lambda_1(\mathbf{vv}_1) + \lambda_2(\mathbf{vv}_2).$$

Так как $l \perp l_1$ и $l \perp l_2$, то $\mathbf{vv}_1 = \mathbf{vv}_2 = 0$, поэтому $\mathbf{vu} = 0$ и $l \perp m$. Теорема доказана.

Условия параллельности и перпендикулярности прямой плоскости

Векторное задание прямых и плоскостей позволяет легко определять их параллельность и перпендикулярность. Пусть даны прямая l с направляющим вектором $\mathbf{v}$ и плоскость α с нормалью $\mathbf{n}$ (рис. 221).

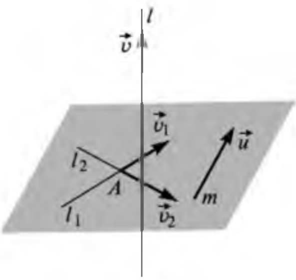


Рис. 220

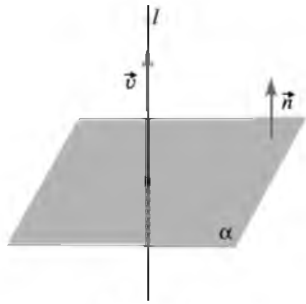


Рис. 221

Прямая l перпендикулярна плоскости α — это означает, что ее направляющий вектор $\mathbf{v}$ коллинеарен нормали $\mathbf{n}$:

$$l \perp \alpha \Leftrightarrow \mathbf{v} \parallel \mathbf{n}.$$

Если прямая l лежит в плоскости α, то ее направляющий вектор $\mathbf{v}$ ортогонален нормали $\mathbf{n}$. Если прямая l параллельна плоскости α (рис. 222), то в этой плоскости всегда есть прямая l', параллельная l. Направляющие векторы $\mathbf{v}$ и $\mathbf{v}'$ прямых l и l' коллинеарны, поэтому если $\mathbf{v}' \perp \mathbf{n}$, то и $\mathbf{v} \perp \mathbf{n}$ (рис. 223). Итак,

$$l \in \alpha \text{ или } l \parallel \alpha \Leftrightarrow \mathbf{v} \perp \mathbf{n}.$$

Выведенные условия позволяют доказывать различные теоремы о параллельности и перпендикулярности прямых и плоскостей.

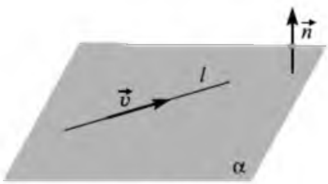


Рис. 222

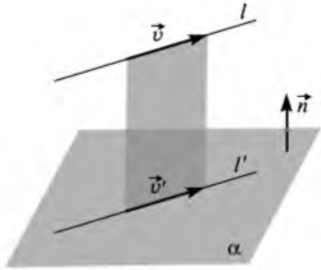


Рис. 223

Пример

Плоскость α и прямая l, перпендикулярные прямой m, параллельны.

Запишем кратко условие примера.

Объекты	*Взаимное расположение*
плоскость α	$\alpha \perp m$;
прямые l, m	$l \perp m$.

Доказать: $l \parallel \alpha$.

Доказательство. Пусть $\mathbf{n}$ — нормаль к плоскости α, $\mathbf{v}$ и $\mathbf{u}$ — направляющие векторы прямых l и m. По условию (рис. 224):

$$\alpha \perp m \Rightarrow \mathbf{n} \parallel \mathbf{u}, \text{ т.е. } \mathbf{u} = \lambda\mathbf{n};$$
$$l \perp m \Rightarrow \mathbf{v} \perp \mathbf{u}, \text{ т.е. } \mathbf{vu} = 0.$$

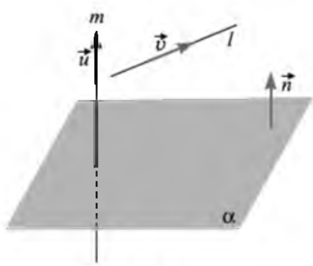


Рис. 224

Подставим первое равенство во второе:

$$\mathbf{vu} = \mathbf{v}\lambda\mathbf{n} = \lambda(\mathbf{vn}) = 0.$$

Так как $\lambda \neq 0$, то $\mathbf{vn} = 0$, т.е. $\mathbf{v} \perp \mathbf{n}$, и, значит, прямая l параллельна плоскости α (или лежит в ней).

Наклонная к плоскости

Если прямая l не параллельна и не перпендикулярна плоскости α, то ее называют наклонной к этой плоскости.

Чтобы определить угол между наклонной и плоскостью, надо спроектировать прямую l на плоскость α. Для этого выбирают точку P на прямой l, отличную от точки P_0 пересечения наклонной и плоскости. Затем, проектируя точку P на плоскость α, получают точку Q. Прямая P_0Q и является проекцией прямой l на плоскость α.

Углом между наклонной и плоскостью называется угол между наклонной и ее проекцией на эту плоскость (рис. 225).

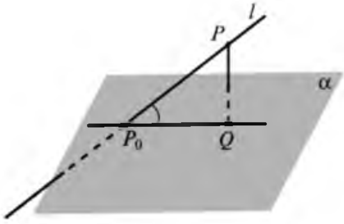


Рис. 225

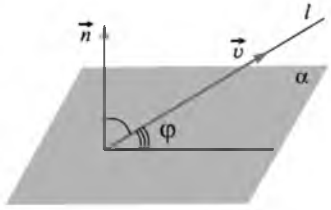


Рис. 226

Пусть **v** — направляющий вектор прямой *l*, а **n** — нормаль к плоскости α. Из рисунка 226 ясно, как связаны между собой угол между прямой *l* и плоскостью α и угол между векторами **v** и **n**. Всегда можно так согласовать направление нормали с направлением вектора **v**, что угол между ними будет острым. Если обозначить острый угол между *l* и α через φ, то сумма угла φ и угла между векторами **v** и **n** равна $\frac{\pi}{2}$. С помощью скалярного произведения можно определить угол φ:

$$\sin\varphi = \sin(\pi / 2 - \angle(\mathbf{v},\mathbf{n})) = \cos\angle(\mathbf{v},\mathbf{n}) = \frac{\mathbf{vn}}{|\mathbf{v}||\mathbf{n}|}.$$

Теорема о трех перпендикулярах

Теорема. Если наклонная к плоскости перпендикулярна какой-нибудь прямой в этой плоскости, то и проекция наклонной перпендикулярна этой прямой. Обратно: если проекция наклонной перпендикулярна прямой в плоскости, то и сама наклонная перпендикулярна этой прямой.

Запишем кратко условие теоремы.

Объекты	*Взаимное расположение*
плоскость α	*l* — наклонная к α;
прямые *l*, *l'* и *m*	*l'* — проекция *l* на α.

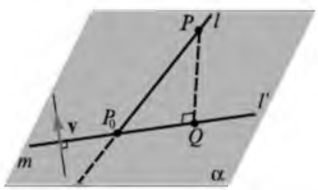


Рис. 227

Доказать: (рис. 227): $m \in \alpha$.

1) $l \perp m \Rightarrow l' \perp m$;

2) $l' \perp m \Rightarrow l \perp m$.

Доказательство. Пусть P_0 — точка пересечения наклонной *l* и плоскости α, $P \neq P_0$ — произвольная точка прямой *l*, *Q* — ее проекция на плоскость α. Тогда вектор

$\overrightarrow{PQ} \perp \alpha$, т.е. является нормалью к плоскости α. Вектор $\overrightarrow{P_0P}$ является направляющим вектором прямой l, а вектор $\overrightarrow{P_0Q}$ — направляющим вектором прямой l'. Эти векторы связаны соотношением $\overrightarrow{P_0P} + \overrightarrow{PQ} = \overrightarrow{P_0Q}$. Кроме того, $\mathbf{v}$ — направляющий вектор прямой m. Докажем первое из двух утверждений теоремы.

Пусть $l \perp m$. Тогда $\overrightarrow{P_0P} \perp \mathbf{v}$. Для проверки того, что $l' \perp m$, вычислим скалярное произведение направляющих векторов этих прямых:

$$\overrightarrow{P_0Q} \cdot \mathbf{v} = (\overrightarrow{P_0P} + \overrightarrow{PQ}) \cdot \mathbf{v} = \overrightarrow{P_0P} \cdot \mathbf{v} + \overrightarrow{PQ} \cdot \mathbf{v}.$$

Первое слагаемое равно нулю по условию, а второе — из-за того, что $\mathbf{v}$, как вектор в плоскости α, перпендикулярен нормали $\overrightarrow{PQ}$ к этой плоскости.

Второе утверждение доказывается аналогично. Проведите доказательство самостоятельно.

Теорема о трех перпендикулярах показывает, что проектирование точки на прямую можно осуществить так: сначала проектируют точку на плоскость, содержащую эту прямую, а затем полученную точку, оставаясь в плоскости, проектируют на прямую (рис. 228). Это подтверждает еще один способ нахождения координат точки в декартовой системе координат: чтобы спроектировать точку P, например на ось x, можно сначала ее спроектировать на плоскость xOy, получить точку P_{xy}, а затем уже в этой плоскости спроектировать точку P_{xy} на ось и получить точку P_x (рис. 229).

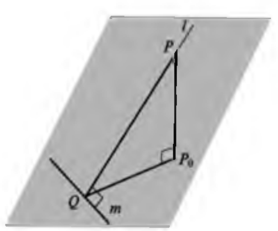


Рис. 228

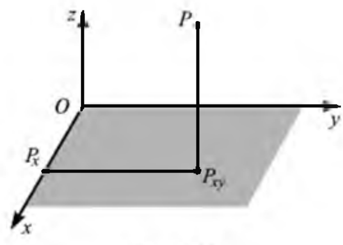


Рис. 229

Ортогональное проектирование

При построении проекции наклонной мы использовали проектирование на плоскость. Рассмотрим это преобразование более подробно.

Пусть дана плоскость α. Ортогональной проекцией точки P на плоскость α называется такая точка Q плоскости α, что прямая PQ перпендикулярна α (рис. 230). Если точка P лежит в плоскости α, то она и считается своей проекцией.

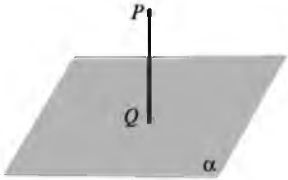


Рис. 230

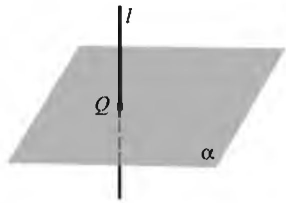


Рис. 231

Пусть даны плоскость α и точка $Q \in \alpha$. Проведем через точку Q прямую l, перпендикулярную плоскости α. Тогда точки прямой l — это все точки, проектирующиеся в точку Q (рис. 231).

При проектировании на прямую l все пространство представляется как бы расслоенным на параллельные плоскости, перпендикулярные прямой l, а при проектировании на плоскость α пространство расслаивается на параллельные прямые, перпендикулярные плоскости α (рис. 232).

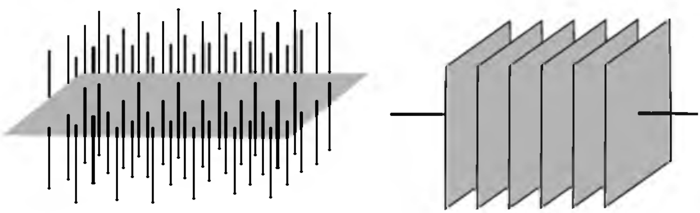


Рис. 232

Ортогональное проектирование на плоскость широко применяется в техническом черчении — детали на чертеже изображаются своими ортогональными проекциями на разные плоскости.

Рассмотрим прямоугольный треугольник ABC, у которого угол B — прямой, сторона BC лежит в плоскости α, а вершина A находится вне ее. Спроектируем вершину A на плоскость α, получим точку O. Полученная конфигурация часто встречается в различных задачах (рис. 233).

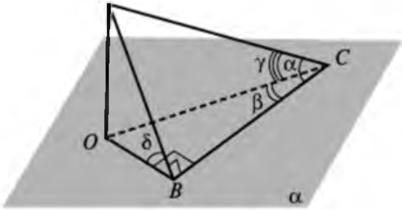


Рис. 233

Сначала заметим, что OB является проекцией наклонной AB и поэтому $OB \perp BC$, т.е. треугольник OBC тоже прямоугольный.

Вычислим соотношения между элементами треугольника ABC и его проекции. Для удобства введем следующие обозначения:

$|AC| = b$, $|CB| = a$, $|OC| = d$, $\angle ACB = \alpha$, $\angle OCB = \beta$, $\angle ACO = \gamma$, $\angle ABO = \delta$.

Вычислим косинусы углов α, β, γ:

$$\cos\alpha = a/b,\ \cos\beta = a/d,\ \cos\gamma = d/b.$$

Перемножая последние два равенства, получаем первое:

$$\cos\alpha = \cos\beta\cos\gamma.$$

Следствие

Угол γ меньше угла α, т.е. угол между наклонной и ее проекцией меньше любого угла, образуемого этой наклонной с прямой на плоскости.

Действительно, так как $0 < \cos\beta < 1$, то $\cos\beta\cos\gamma < \cos\gamma$, т.е. $\cos\alpha < \cos\gamma$, а тогда $\alpha > \gamma$, так как косинус убывает в первой четверти. Аналогично доказывается, что $\alpha > \beta$.

Вычислим площади треугольников ABC и OBC:

$$S_{ABC} = {}^1/_2\, a\,|AB|,\ S_{OBC} = {}^1/_2\, a\,|OB|.$$

Видим, что отношение площадей $\dfrac{S_{OBC}}{S_{ABC}}$ равно отношению

$$\frac{|OB|}{|OA|} = \cos\delta, \text{ т.е. } S_{OBC} = S_{ABC}\cos\delta.$$

Двугранные углы

Линейный угол двугранного угла

Две непараллельные плоскости пересекаются по прямой и делят все пространство на четыре части, называемые двугранными углами.

Двугранный угол — это фигура, ограниченная двумя полуплоскостями с общим ребром (рис. 234).

Четыре двугранных угла, образованных двумя пересекающимися плоскостями, распадаются на две пары равных между собой двугранных углов, что аналогично равенству вертикальных углов, образующихся при пересечении двух прямых (рис. 235).

Двугранные углы можно измерять с помощью линейных углов.

Рис. 234

Рис. 235

Построение линейного угла

Возьмем на ребре двугранного угла точку и восстановим в гранях по перпендикуляру к этому ребру. Угол, образованный построенными перпендикулярами, называется *линейным углом двугранного угла* (рис. 236).

Рис. 236

Рис. 237

В качестве вершины линейного угла можно взять любую точку на ребре двугранного угла. Все такие углы будут равны между собой как углы с соответственно параллельными сторонами (рис. 237).

Равные двугранные углы имеют равные линейные углы, и, наоборот, двугранные углы с равными линейными углами равны между собой.

Плоскость, проходящая через стороны линейного угла, перпендикулярна ребру двугранного угла, так как это ребро перпендикулярно двум пересекающимся прямым (сторонам угла), лежащим в этой плоскости.

Рис. 238

Обратно: если мы проведем плоскость, перпендикулярную ребру двугранного угла, то в сечении получится линейный угол (рис. 238). Действительно, ребро перпендикулярно плоскости сечения, следовательно, перпендикулярно прямым, по которым эта плоскость пересекает грани двугранного угла. Эти прямые образуют, по определению, линейный угол.

Перпендикулярные плоскости

Две плоскости называются перпендикулярными, если, пересекаясь, они образуют равные двугранные углы.

Перпендикулярные плоскости образуют двугранные углы по 90° (рис. 239).

Теорема (признак перпендикулярности двух плоскостей). Если плоскость проходит через прямую, перпендикулярную другой плоскости, то эти плоскости перпендикулярны (рис. 240).

Рис. 239

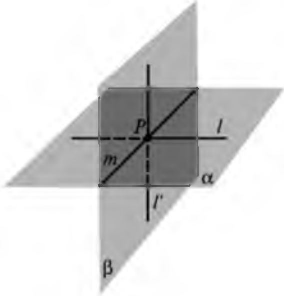


Рис. 240

Запишем кратко условие теоремы.

Объекты	*Взаимное расположение*
плоскости α, β	$l \in \alpha$;
прямая l	$l \perp \beta$.

Доказать: $\alpha \perp \beta$.

Доказательство. Пусть P — точка пересечения прямой l с плоскостью β. Плоскости α и β пересекаются ($P \in \alpha$, $P \in \beta$) по прямой m. Прямые l и m перпендикулярны (определение перпендикуляра к плоскости). Проведем в плоскости β через точку P прямую l', перпендикулярную m. Прямые l и l' образуют линейный угол двугранного угла с ребром m, так как обе они перпендикулярны ребру. Но $l \perp l'$ (по определению перпендикуляра к плоскости). Следовательно, линейный угол прямой, т.е. плоскости α и β перпендикулярны. Теорема доказана.

Если плоскости заданы векторными уравнениями, то их перпендикулярность легко проверить, вычисляя скалярное произведение их нормалей. Очевидно, что если плоскости перпендикулярны, то и нормали к ним перпендикулярны и их скалярное произведение равно нулю.

Это позволяет доказывать новые теоремы о перпендикулярности прямых и плоскостей.

Пример

Если прямая l параллельна плоскости α и перпендикулярна плоскости β, то плоскости α и β перпендикулярны (рис. 241).

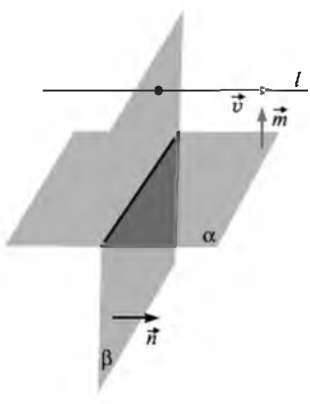


Рис. 241

Запишем кратко условие примера.

Объекты	*Взаимное расположение*
прямая l	$l \parallel \alpha$;
плоскости α, β	$l \perp \beta$.

Доказать: $\alpha \perp \beta$.

Пусть $\mathbf{v}$ — направляющий вектор прямой l, $\mathbf{m}$ и $\mathbf{n}$ — нормали к плоскостям α и β. По условию,

$$l \parallel \alpha \Rightarrow \mathbf{v} \perp m, \; l \perp \beta \Rightarrow \mathbf{v} \parallel \mathbf{n}.$$

Если $\mathbf{n} \parallel \mathbf{v}$, а $\mathbf{v} \perp \mathbf{m}$, то и $\mathbf{n} \perp \mathbf{m}$, т.е. $\alpha \perp \beta$.

Угол между плоскостями

Пусть даны две пересекающиеся плоскости. Они образуют между собой два угла, сумма которых равна 180°. Выберем один из двугранных углов, образованных этими плоскостями. Обозначим его φ. Проведем к плоскостям нормали так, чтобы они были направлены внутрь двугранного угла. Из рисунка 242 ясно, что сумма угла φ и угла между нормалями тоже равна 180°. Поэтому вычисление угла между плоскостями можно заменить вычислением угла между нормалями.

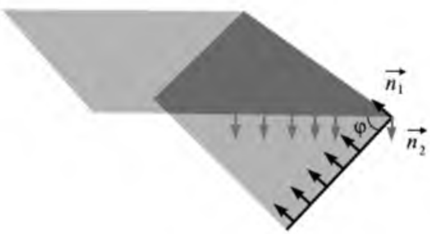


Рис. 242

Пусть $\mathbf{n}_1$ и $\mathbf{n}_2$ — нормали к данным плоскостям. Косинус угла между двумя векторами определяется по известной формуле:

$$\cos\varphi = \frac{\mathbf{n}_1\mathbf{n}_2}{|\mathbf{n}_1||\mathbf{n}_2|}.$$

По этой же формуле находится и угол между плоскостями. При этом надо помнить, что если скалярное произведение получится отрицательным, то это означает, что найден косинус большего (тупого) угла из двух углов, образованных плоскостями.

Контрольные вопросы и задания

1. Дайте определение прямой, перпендикулярной плоскости.
2. Что такое нормаль к плоскости?
3. Каково векторное уравнение плоскости?

4. Как записать уравнение плоскости в координатах?
5. Какой смысл имеют коэффициенты при переменных в координатном уравнении плоскости?
6. Сформулируйте признак перпендикулярности прямой и плоскости.
7. Сформулируйте условия параллельности и перпендикулярности прямой и плоскости.
8. Что называется углом между прямой и плоскостью?
9. В чем состоит теорема о трех перпендикулярах?
10. Каким свойством обладает угол между наклонной и ее проекцией?
11. Что такое двугранный угол?
12. Что такое линейный угол двугранного угла?
13. Какие плоскости называются перпендикулярными?
14. Сформулируйте признак перпендикулярности двух плоскостей.
15. Как можно вычислить угол между двумя плоскостями?

ГЛАВА 6

ТЕЛА И ПОВЕРХНОСТИ В ПРОСТРАНСТВЕ

§ 25. Цилиндры и конусы

Прямой круговой цилиндр

Форма тел обычно сложна. Однако, как правило, тело можно разбить на части, имеющие более простую форму. На рисунке 243 изображен древнегреческий храм — акрополь. В его структуре ясно выделяются цилиндрические колонны, основной корпус в виде коробки — параллелепипеда, увенчанного треугольным фронтоном. Как из тел простой формы — параллелепипедов, призм, пирамид, шаров, цилиндров — создать прекрасное здание, — это секрет искусства, архитектуры. Однако основа этого секрета кроется в геометрии, в использовании различных свойств симметрии тел и их сочетаний.

Рис. 243

Нашей задачей является изучение геометрических свойств простейших тел. Эти тела — многогранники, тела вращения, конусы, цилиндры.

Цилиндр — одно из самых распространенных тел. Простейшим цилиндром является прямой круговой цилиндр.

Прямым круговым цилиндром называется тело, образованное вращением прямоугольника вокруг одной из его сторон (рис. 244).

В цилиндре можно выделить два основания, лежащие в параллельных плоскостях и представляющие собой круги одинакового радиуса. Прямая, проходящая через центры оснований, называется *осью цилиндра*.

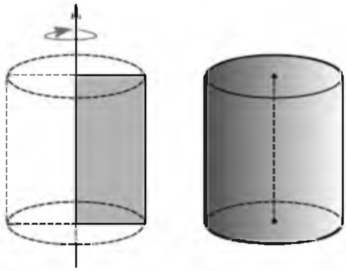


Рис. 244

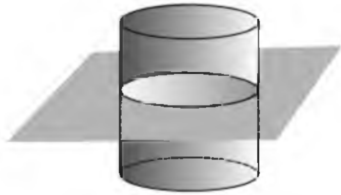


Рис. 245

Границу цилиндра составляют два основания и боковая поверхность цилиндра. Ее можно представлять себе образованной вращением отрезка вокруг оси цилиндра.

Сечением цилиндра плоскостью, параллельной основанию, будет круг того же радиуса, что и основание (рис. 245).

Высота цилиндра — это расстояние между основаниями, т.е. длина отрезка оси между центрами оснований. Обычно цилиндр задают двумя числовыми данными: радиусом основания и высотой.

Боковую поверхность цилиндра можно развернуть на плоскость. Если представить себе эту поверхность сделанной из бумаги и разрезать ее параллельно оси цилиндра, то она развернется в форме прямоугольника, одна из сторон которого равна высоте цилиндра, а другая — длине окружности основания (рис. 246).

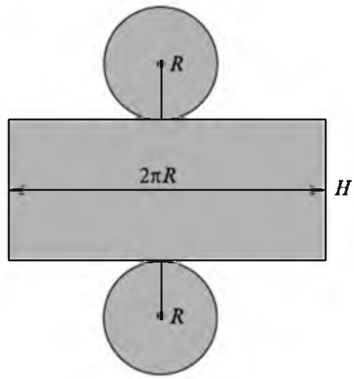


Рис. 246

Прямой круговой конус

Коническая форма так же, как и цилиндрическая, широко распространена в окружающем нас мире. В форме конуса делают колпаки и кульки, различные втулки и емкости, коническую форму принимают насыпанная куча песка (рис. 247) и воронка от взрыва. Фонарь освещает часть пространства в форме конуса.

Прямым круговым конусом называется тело, образованное вращением прямоугольного треугольника вокруг одного из его катетов (рис. 248).

Рис. 247

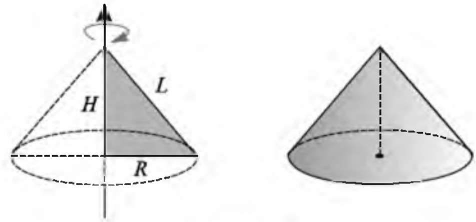


Рис. 248

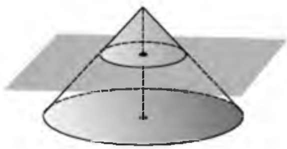


Рис. 249

В конусе можно выделить основание в форме круга и вершину. Прямая, проходящая через вершину и центр основания, называется осью конуса.

Боковая поверхность конуса образована вращением вокруг оси отрезка, один конец которого совпадает с вершиной конуса.

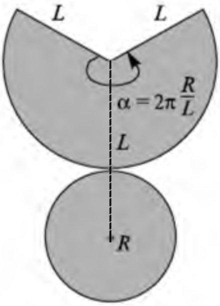


Рис. 250

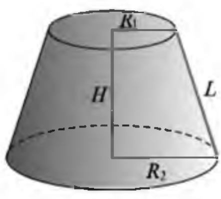


Рис. 251

Сечением конуса плоскостью, параллельной основанию, будет круг (рис. 249).

Высота конуса — это расстояние от его вершины до основания, т.е. длина отрезка оси, соединяющего вершину и центр основания. Конус можно задать двумя числовыми данными: радиусом основания и высотой.

Радиус основания и высота — это катеты того прямоугольного треугольника, вращением которого был образован конус. Гипотенуза этого треугольника называется *образующей конуса*.

Так же, как и у цилиндра, боковую поверхность конуса можно развернуть на плоскость. Если разрез сделать по образующей, то разверткой будет круговой сектор (рис. 250).

Если в конусе провести сечение, параллельное основанию, то между основанием и плоскостью сечения получится тело, называемое усеченным конусом (рис. 251). Оставшаяся часть конуса (от вершины до плоскости сечения) будет также конусом, подобным исходному. Коэффициент подобия равен отношению высот (или отношению радиусов оснований).

Тела вращения

Цилиндр и конус являются простейшими примерами тел вращения. Тела вращения более общего вида получаются при вращении плоской фигуры вокруг не пересекающей ее оси. Издавна человек изготовлял с помощью гончарного круга домашнюю утварь в форме тел вращения. Аналогичный принцип применяется теперь при обработке металлов на различных станках, вращающих заготовку. Наблюдая некоторые рукотворные явления, имитируюшие природные (водовороты, смерчи и др.), можно увидеть появление тел и поверхностей вращения (рис. 252, 253).

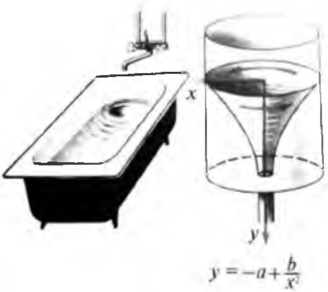


Рис. 252

Рис. 253

Кроме цилиндров и конусов в технике часто встречаются такие тела, как параболоид вращения, полученный при вращении параболы вокруг ее оси (рис. 254). Он используется при изготовлении различных фокусирующих зеркальных устройств (рис. 255). Эффект его применения описан в известном романе А.Н. Толстого «Гиперболоид инженера Гарина». При вращении окружности вокруг не пересекающей ее оси получается тело в форме баранки. В математике это тело называют тором (рис. 256).

Общее представление о цилиндре и конусе

Прямой круговой цилиндр и прямой круговой конус были нами определены с помощью операции вращения. Это привело нас к одному обобщению — к понятию тела вращения. Однако цилиндр и конус можно получить иначе, используя другие свойства этих тел.

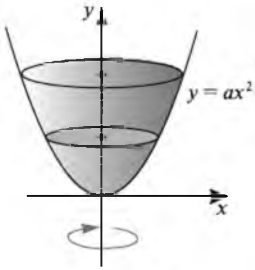


Рис. 254

Рис. 255

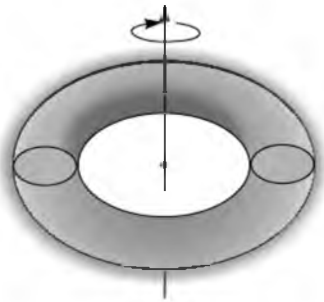


Рис. 256

Обычный, т.е. прямой, круговой цилиндр можно получить так: взять круг и сделать его параллельный перенос на вектор, ортогональный плоскости этого круга; затем соединить отрезком каждую точку исходного круга с соответствующей точкой нового, передвинутого круга. Цилиндр получится как объединение всех этих отрезков. Это построение можно обобщить: выбрать в качестве исходной фигуры не обязательно круг, а произвольную плоскую фигуру, и совершать параллельный перенос на произвольный вектор.

Пусть даны плоская фигура *F* и вектор а, не лежащий в плоскости фигуры. Цилиндром называется тело, образованное отрезками, соединяющими точки фигуры *F* с точками, получающимися из них при параллельном переносе на вектор а (рис. 257).

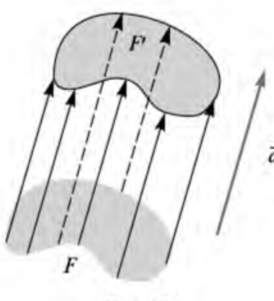


Рис. 257

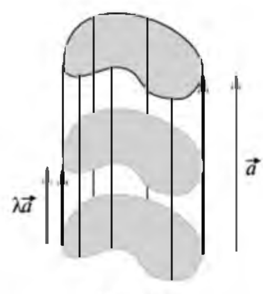


Рис. 258

Фигура *F* и равная ей фигура, получающаяся при параллельном переносе *F* на вектор а, называются *основаниями цилиндра*. Расстояние между плоскостями оснований называется *высотой цилиндра*. Если вектор а ортогонален плоскости основания, то цилиндр прямой (рис. 258).

В сечении цилиндра плоскостью, параллельной основанию, получается фигура, равная основанию. Вообще, если цилиндр получен с помощью фигуры F и параллельного переноса на вектор **a**, то сечение этого цилиндра плоскостями, параллельными основанию, можно получать при параллельном переносе фигуры F на векторы вида $\lambda\mathbf{a}$, где $0 \le \lambda \le 1$.

Обычный, т.е. прямой, круговой конус можно получить так: взять круг, выбрать точку, проектирующуюся в центр этого круга, и соединить ее со всеми точками круга. Конус получится как объединение всех этих отрезков. Обобщим это построение, взяв в качестве основания произвольную плоскую фигуру, а в качестве вершины — произвольную точку.

Пусть даны плоская фигура F и точка O вне плоскости фигуры. Конусом называется тело, образованное отрезками, соединяющими точку O со всеми точками фигуры F (рис. 259).

Точка O называется *вершиной конуса*, фигура F — его *основанием*, расстояние от вершины до плоскости основания — *высотой*.

В сечении конуса плоскостью, параллельной основанию, получается фигура, подобная фигуре F. Коэффициент подобия равен отношению расстояния от вершины до плоскости сечения к высоте конуса. Преобразование подобия осуществляется как раз с помощью тех отрезков, которые соединяют вершину конуса с точками основания.

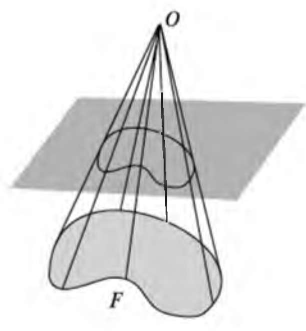


Рис. 259

Контрольные вопросы и задания

1. Какое тело называется прямым круговым цилиндром?
2. Как можно построить прямой круговой конус?
3. Приведите примеры тел вращения.
4. Дайте общее определение цилиндра.
5. Как в общем виде строится конус?

§ 26. Шар и сфера

Определения

Сферой радиуса R (где $R > 0$) с центром в точке O называется множество точек пространства, удаленных от точки O на расстоянии R (рис. 260).

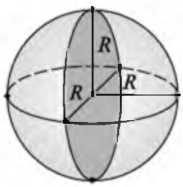


Рис. 260

Сфера разбивает пространство на две части: внутреннюю, составленную из точек, удаленных от точки O на расстояние, меньшее R, и внешнюю, составленную из точек, удаленных от точки O на расстояние, большее R. Внутренняя часть пространства вместе со сферой называется *шаром радиуса R* с центром в точке O. Таким образом, шар — множество точек пространства, удаленных от точки O на расстояние, меньшее или равное R.

Сфера и шар радиуса R с центром в точке O получаются при вращении соответственно окружности и круга вокруг диаметра.

Отрезки, соединяющие точки сферы с центром, называются *радиусами сферы*. Они же называются радиусами соответствующего шара. *Диаметрами шара и сферы* называются отрезки, по которым шар пересекается с прямыми, проходящими через центр.

Этими же словами — «радиус» и «диаметр» — обозначают не только указанные отрезки, но и их длины.

Сечение шара плоскостью

Теорема. Пусть даны шар радиуса R и плоскость, расстояние которой до центра шара равно d.
1. Если $d > R$, то у шара и плоскости нет общих точек (рис. 261, *а*).
2. Если $d = R$, то у шара и плоскости одна общая точка — проекция центра шара на плоскость (рис. 261, *б*).
3. Если $d < R$, то пересечение шара и плоскости является кругом радиуса $r = \sqrt{R^2 - d^2}$ (рис. 261, *в*).

Доказательство. Даны шар S с центром O и плоскостью α. Спроектируем точку O на плоскость α. Получим точку $P \in \alpha$. Дано, что $|OP| = d$ (рис. 262).

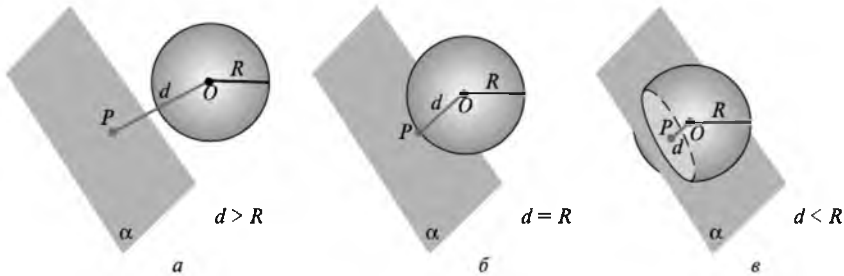


Рис. 261

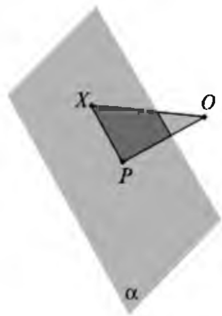


Рис. 262

Пусть X — произвольная точка плоскости α. Вычислим расстояние от нее до центра шара: $|OX|^2 = |PX|^2 + d^2$.

1. $d > R$. Тогда $|OX|^2 > |PX|^2 + R^2 > R^2$, т.е. $|OX| > R$. Это означает, что расстояние от точки X до центра шара больше радиуса, значит, точка X лежит вне шара. Это верно для любой точки X плоскости α. Следовательно, все точки плоскости α лежат вне шара и шар с плоскостью не пересекаются.
2. $d = R$ и точка X принадлежат шару. Тогда $|OX| \le R$. Подставив $d = R$ и $|OX| \le R$ в формулу $|OX|^2 = |PX|^2 + d^2$, получим $|PX|^2 + R^2 \le R^2$, т.е. $|PX|^2 \le 0$.

Это возможно лишь в том случае, когда $|PX| = 0$, т.е. когда точки P и X совпадают. Вспомним, что P — однозначно определенная точка, проекция центра шара на плоскость α. Итак, мы доказали, что общая точка у шара и плоскости одна — проекция центра шара на плоскость.

3. $d < R$ и точка X принадлежат шару. Тогда $|OX| \le R$. Введем обозначения: $r = \sqrt{R^2 - d^2}$. Из $\Delta\, OPX$ имеем $|PX|^2 = |OX|^2 - d^2 \le R^2 - d^2 = r^2$, т.е. $|PX| \le r$. Мы получили, что расстояние от точки X до точки P не превосходит r. Это означает, что точка X лежит в круге с центром в точке P и радиусом r. Заметим, что все точки этого круга лежат в сечении шара плоскостью. Действительно, если $X \in \alpha$ и лежит в круге с центром P и радиусом r, то $|PX| \le r$. Тогда $|OX|^2 = |PX|^2 + d^2 = |PX|^2 + R^2 - r^2 \le r^2 + R^2 - r^2 = R^2$, т.е. $|OX| \le R$, откуда следует, что $X \in S$, т.е. $X \in S \cap \alpha$. Теорема доказана.

Аналогично можно рассмотреть пересечение плоскости и сферы. Сформулируйте самостоятельно теорему о пересечении плоскости и сферы.

В сечении шара радиуса R плоскостью (если они пересекаются) получается круг некоторого радиуса $r = \sqrt{R^2 - d^2}$, где d — расстояние от центра шара до плоскости. Так как это расстояние может изменяться от 0 до R, то и r как функция от d может изменяться в этих же пределах.

При $d = 0$ (в том случае, когда плоскость проходит через центр шара) в сечении получаются круги максимального радиуса R. Они называются *большими кругами*.

Касательная плоскость

Теорема о пересечении плоскости и шара показывает, что возможны три случая расположения шара и плоскости: шар и плоскость не имеют общих точек, шар и плоскость имеют одну общую точку, общие точки шара и плоскости заполняют круг.

Плоскость называется касательной к шару, если она имеет с ним одну общую точку (рис. 263).

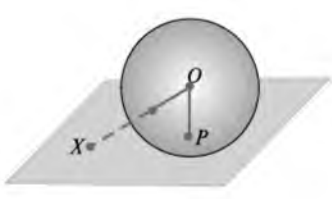


Рис. 263

Аналогично определяется касательная плоскость к сфере.

При доказательстве теоремы о пересечении плоскости и шара мы выяснили, что одна общая точка у них будет в том случае, когда радиус шара равен расстоянию от центра шара до плоскости. При этом единственная общая точка (точка касания) будет проекцией центра шара на плоскость. Это свойство можно сформулировать следующим образом.

Теорема. Пусть P — общая точка плоскости α и шара. Если радиус шара, проведенный в точку P, перпендикулярен плоскости α, то эта плоскость α является касательной к шару, а точка P — точка касания.

Доказательство. Пусть O — центр шара. По условию, $OP \perp \alpha$ и $|OP| = R$, где R — радиус шара. Пусть X — любая другая точка плоскости α. Тогда $|OX| > |OP|$, так как перпендикуляр к плоскости короче наклонной. Следовательно, у шара и плоскости α есть единственная общая точка P и, по определению, плоскость α касается шара. Теорема доказана.

Шар нельзя развернуть на плоскость. Каждый из нас сталкивался с тем, что бумагой, не сминая ее, нельзя оклеить шар (рис. 264), как нельзя скроить круглую шляпу из нерастяжимой материи. Это замечательное свойство шара, отличающее его от конусов и цилиндров, кажется нам вполне очевидным, однако получить его математическое доказательство очень трудно.

Рис. 264

Контрольные вопросы и задания

1. Что такое сфера?
2. В чем разница между сферой и шаром?
3. Что получается в сечении шара плоскостью?
4. Как определяется плоскость, касательная к шару?
5. Как построить касательную плоскость к шару?

§ 27. Призмы и пирамиды

Призмы

Призмой называется цилиндр, в основании которого лежит многоугольник (рис. 265).

По определению цилиндра получаем следующий способ построения призмы. Берем многоугольник F и вектор **a**, не лежащий в плоскости этого многоугольника. Каждую точку многоугольника F переносим на вектор **a**. Соединяем полученные точки отрезками и объединяем вместе все отрезки.

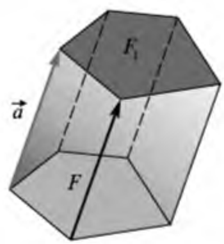


Рис. 265

Исходный многоугольник F удобно называть ***нижним основанием призмы***, а многоугольник, полученный из F параллельным переносом на вектор **a**, — ***верхним основанием***. Отрезки, соединяющие соответствующие вершины верхнего и нижнего оснований, называются ***боковыми ребрами призмы***.

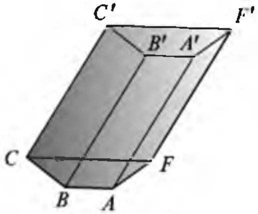


Рис. 266

Теорема. Все боковые ребра призмы параллельны друг другу и равны между собой (рис. 266).

Доказательство. Пусть A и B — две вершины нижнего основания призмы, A' и B' — соответствующие им вершины верхнего основания. Так как точки A' и B' получены из точек A и B параллельным переносом на один и тот же вектор **a**, то $\overline{AA'} = \overline{BB'} = \mathbf{a}$, а это и означает, что $AA' \parallel BB'$ и $|AA'| = |BB'|$, что и требовалось доказать.

Следствие. Боковые грани призмы являются параллелограммами.

Действительно, они являются четырехугольниками, имеющими пару равных и параллельных сторон.

Введем определения.

Треугольная, четырехугольная, ..., *n*-угольная призма — в основании призмы лежит треугольник, четырехугольник, ..., *n*-угольник (рис. 267 и 268). **Прямая призма — боковые ребра призмы перпендикулярны основанию** (рис. 269).

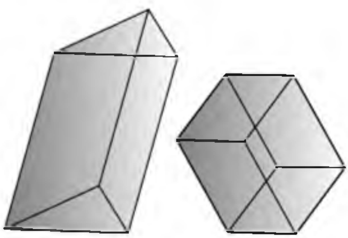


Рис. 267

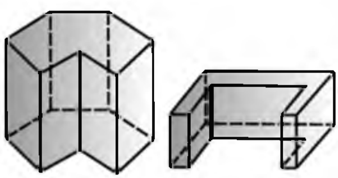


Рис. 268

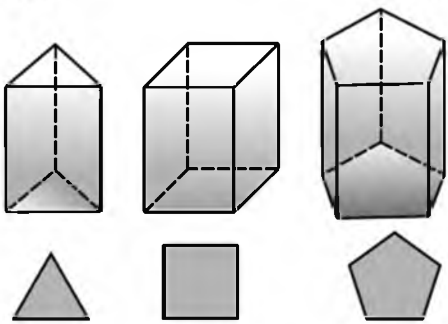


Рис. 269

Определение прямой призмы совпадает с общим определением прямого цилиндра.

Наклонная призма, т.е. не прямая, ее боковые ребра не перпендикулярны основанию.
Правильная призма — прямая и в ее основании лежит правильный многоугольник.

Диагональ призмы — это отрезок, соединяющий вершины двух оснований призмы и лежащий внутри ее (не лежащий в боковой грани) (рис. 270).

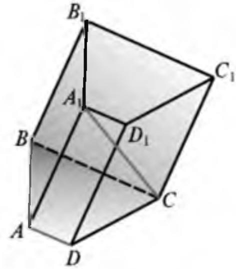


Рис. 270

Пирамиды

Пирамидой называется конус, в основании которого лежит многоугольник.

Вспоминая определение конуса, получаем следующий способ построения пирамиды. Берем многоугольник F и точку O, не лежащую в его плоскости. Соединяем точку O с каждой точкой многоугольника F и объединяем все построенные отрезки (рис. 271).

Отрезки, соединяющие вершину пирамиды с вершинами основания, называются *боковыми ребрами пирамиды*.

Введем определения.

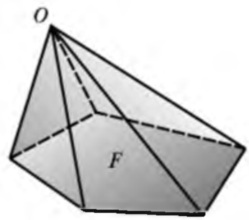


Рис. 271

Треугольная, четырехугольная, ..., *n*-угольная пирамида — в основании лежит треугольник, четырехугольник, ..., *n*-угольник (рис. 272). **Треугольную пирамиду еще иначе называют тетраэдром.**
Правильная пирамида — в основании лежит правильный многоугольник, а вершина пирамиды проектируется в центр этого многоугольника (рис. 273).
Усеченная пирамида — часть пирамиды, отсекаемая плоскостью, параллельной основанию.

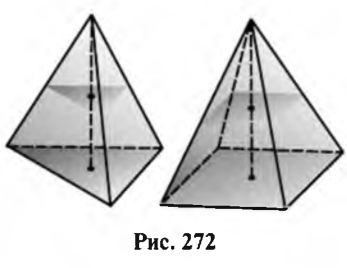


Рис. 272

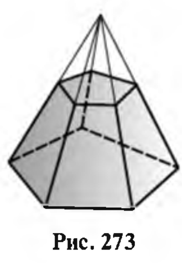


Рис. 273

Напомним, что, как и для всякого конуса, в сечении пирамиды плоскостью, параллельной основанию, получается фигура, подобная основанию. Поэтому, например, в таких сечениях правильной пирамиды будут получаться правильные многоугольники.

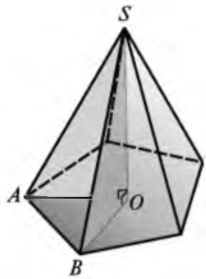


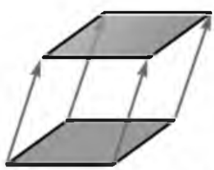
Рис. 274

Теорема. В правильной пирамиде боковые ребра равны и одинаково наклонены к плоскости основания (рис. 274).

Доказательство. Пусть S вершина пирамиды, A и B две произвольные вершины основания, O его центр. По условию, отрезок SO перпендикулярен основанию (вершина правильной пирамиды проектируется в его центр); $|OA| = |OB|$ как радиусы круга, описанного около основания. Следовательно, прямоугольные треугольники OSA и OSB равны, так как имеют по два равных катета. Отсюда вытекает, что $|SA| = |SB|$ и $\angle SAO = \angle SBO$, что и требовалось доказать.

Параллелепипеды

Параллелепипедом называется призма, в основании которой лежит параллелограмм (рис. 275).



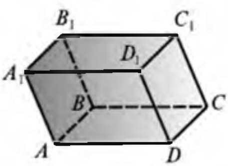


Рис. 275

Боковые грани параллелепипеда, как у всякой призмы, являются параллелограммами. Поэтому в отличие от произвольной призмы у параллелепипеда нельзя однозначно выделить основания. В качестве оснований можно взять любую из трех пар противоположных граней.

Если продолжить все грани параллелепипеда, то они образуют три пары параллельных плоскостей. Вспомним, параллелограмм можно получить, рассматривая пересечение двух пар параллельных прямых. Аналогично, и параллелепипед получается с помощью пересечения трех пар параллельных плоскостей (рис. 276).

В параллелепипеде можно провести четыре диагонали.

Теорема. Четыре диагонали параллелепипеда пересекаются в одной точке, являющейся серединой каждой диагонали (рис. 277).

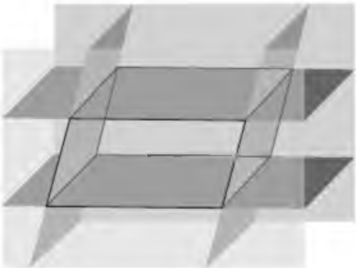


Рис. 276

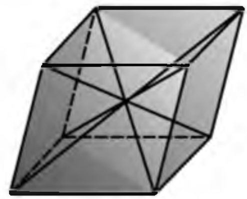


Рис. 277

Доказательство. Возьмем две любые диагонали. Они лежат в одном из диагональных сечений. Это сечение является параллелограммом. Выбранные диагонали параллелепипеда будут диагоналями этого параллелограмма. По свойству параллелограмма диагонали в точке пересечения делятся пополам.

Это верно для любой пары диагоналей, т.е. середины любых двух диагоналей совпадают, значит, они совпадают у всех четырех диагоналей. Эта точка и есть точка пересечения всех диагоналей. Теорема доказана.

Параллелепипед, у которого все грани — прямоугольники, называется прямоугольным (рис. 278).

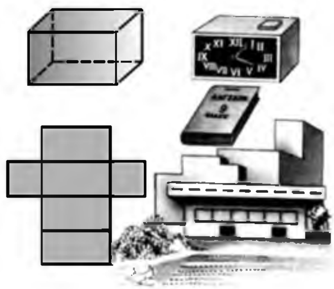


Рис. 278

Пространственная теорема Пифагора

Обычная формулировка *теоремы Пифагора* хорошо известна: в прямоугольном треугольнике квадрат гипотенузы равен сумме квадратов катетов (рис. 279):

$$a^2 + b^2 = d^2.$$

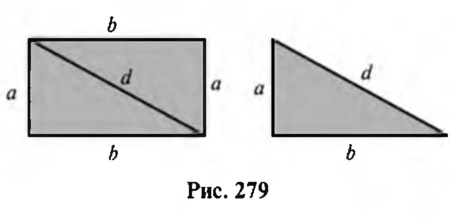


Рис. 279

Вместо прямоугольного треугольника можно рассматривать прямоугольник и дать теореме Пифагора равносильную формулировку.

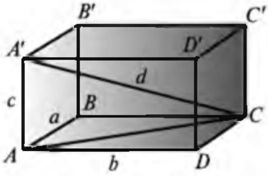


Рис. 280

В прямоугольнике квадрат диагонали равен сумме квадратов сторон, исходящих из одной вершины (рис. 280).

В такой формулировке теорема Пифагора обобщается на пространственный случай. Прямоугольник теперь заменяется прямоугольным параллелепипедом.

Теорема. Квадрат диагонали прямоугольного параллелепипеда равен сумме квадратов ребер, исходящих из одной вершины (рис. 280).

$$d^2 = a^2 + b^2 + c^2.$$

Доказательство. Возьмем диагональ $A'C$. Треугольник $AA'C$ прямоугольный, так как в прямоугольном параллелепипеде боковые ребра перпендикулярны основанию. По теореме Пифагора, на плоскости

$$|A'C|^2 = |AA'|^2 + |AC|^2 = |AA'|^2 + |AD|^2 + |AB|^2,$$

что и требовалось доказать.

Пространственная теорема Пифагора на самом деле мало чем отличается от формулы расстояния между двумя точками в декартовых координатах.

Пусть $M_1(x_1; y_1; z_1)$ и $M_2(x_2; y_2; z_2)$ — две точки (рис. 281). Вектор $\overline{M_1M_2}$ имеет координаты $(x_2 - x_1; y_2 - y_1; z_2 - z_1)$. Квадрат его длины равен

$$|M_1M_2|^2 = (x_2 - x_1)^2 + \\ + (y_2 - y_1)^2 + (z_2 - z_1)^2.$$

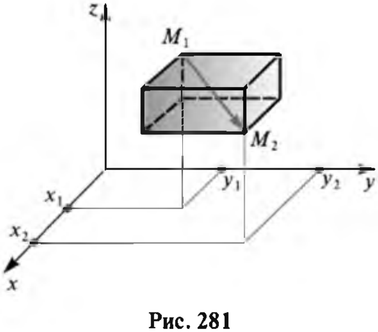


Рис. 281

Отрезок M_1M_2 можно представить как диагональ прямоугольного параллелепипеда с ребрами, параллельными координатным осям. Длины сторон (измерения) этого параллелепипеда равны $|x_2 - x_1|$, $|y_2 - y_1|$, $|z_2 - z_1|$, откуда видно совпадение двух указанных формул.

Контрольные вопросы и задания

1. Дайте определение призмы.
2. Попробуйте дать непосредственное определение призмы, не используя общее понятие цилиндра.
3. Сформулируйте свойства боковых ребер призмы.
4. Какие виды призм вы знаете?
5. Что такое пирамида?
6. Дайте определение пирамиды, не используя общее понятие конуса.
7. Какие виды пирамид вы знаете?
8. Назовите свойства боковых ребер правильной пирамиды.
9. Что такое параллелепипед?
10. Какие варианты теоремы Пифагора вы знаете?

§ 28. Многогранники

Примеры многогранников

Многогранник — это тело, ограниченное многоугольниками (рис. 282).

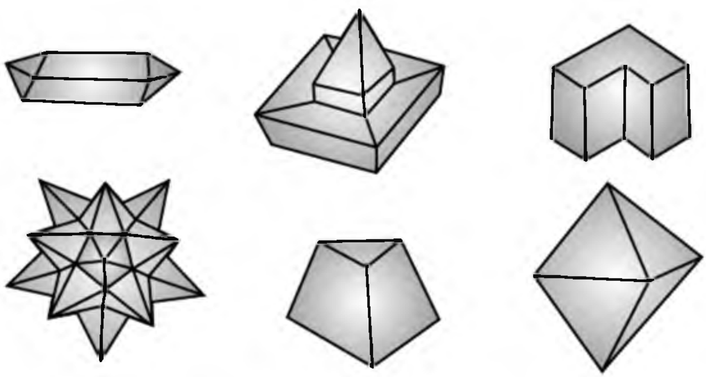


Рис. 282

Данное определение многогранника еще не вполне отвечает обычным представлениям о многограннике. Так, все пространство, кроме точек некоторого куба (т.е. вне точек некоторого куба), тоже можно считать телом, ограниченным многоугольниками. Мы же представляем себе многогранник ограниченным телом.

Рассмотрим тела, ограниченные конечным числом многоугольников, и изучим определенные типы многогранников, которые легко себе представить и модели которых нетрудно изготовить.

Призмы и пирамиды являются простейшими примерами многогранников.

Среди всех тел многогранники выделяются тем, что их поверхность состоит из кусков плоскостей, являющихся многоугольниками. Эти многоугольники называются *гранями*, их стороны — *ребрами*, а вершины — *вершинами многогранника*. Так, треугольная пирамида имеет четыре грани, каждая из которых представляет собой треугольник. Всего у треугольной пирамиды шесть ребер и четыре вершины, в каждой из которых сходится 3 ребра.

У куба 6 граней, 12 ребер и 8 вершин, каждая грань является квадратом. В каждой вершине сходится по три ребра.

Может ли многогранник иметь любое количество граней, ребер и вершин? Оказывается, нет. Давно было замечено, что между количеством граней, ребер и вершин многогранника есть простое соотношение. Попробуйте заметить его сами, посмотрев на табл. 3, в которой указано количество граней, ребер и вершин для шести многогранников (рис. 282).

Таблица 3

Составляющие многогранника

Номер многогранника	Количество		
	граней	ребер	вершин
1	8	14	8
2	13	24	13
3	8	18	12
4	60	90	32
5	5	9	6
6	8	12	6

Легко заметить, что сумма количества граней и количества вершин всегда на 2 больше количества ребер. Это наблюдение верно для любого выпуклого многогранника и составляет содержание знаменитой теоремы, доказанной впервые Леонардом Эйлером (1707—1783).

Теорема Эйлера. Пусть $Г$ обозначает число граней, P — число ребер, B — число вершин выпуклого многогранника. Тогда

$$Г + В = Р + 2.$$

Полное доказательство теоремы Эйлера довольно трудоемко, и мы его приводить не будем. Проверить же эту теорему для частных случаев многогранников достаточно просто.

К многогранникам надо отнести и такие необычные конфигурации, которые изображены на рис. 283. Все они невыпуклые, однако для некоторых из них соотношение Эйлера $\Gamma + B = P + 2$ выполнено, а для некоторых — нет. На самом деле его справедливость связана не с выпуклостью многогранника, а скорее, с другими его свойствами, например наличием или отсутствием «дырок».

Рис. 283

Правильные многогранники

Куб — параллелепипед, у которого все грани квадраты. Все ребра куба имеют одинаковую длину, поскольку все стороны квадрата имеют одинаковую длину. Ребра, выходящие из одной вершины, попарно перпендикулярны. Соседние грани куба перпендикулярны друг другу, поскольку каждая из них содержит ребро, перпендикулярное другой грани.

Куб имеет девять плоскостей симметрии (рис. 284), девять осей симметрии (рис. 285) и центр симметрии (рис. 286). Кроме того, при повороте на углы, кратные 90°, вокруг осей a_1, a_2, a_3 и при повороте на углы, кратные 120°, вокруг диагоналей куба куб в целом сохраняет свое положение (при этом некоторые вершины, ребра и грани, конечно, меняются местами). Всего существует 48 перемещений, включая 24 поворота, при которых куб сохраняет свое положение. С помощью поворотов можно осуществить перестановку диагоналей куба. В частности, эти повороты позволяют поменять местами любые два трехгранных угла куба. И с этой точки зрения все трехгранные углы куба одинаковые.

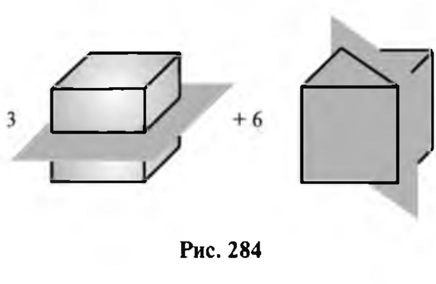


Рис. 284

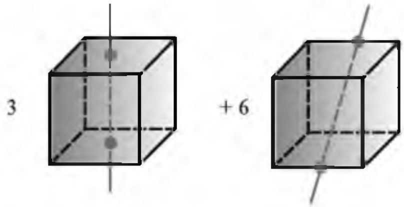


Рис. 285

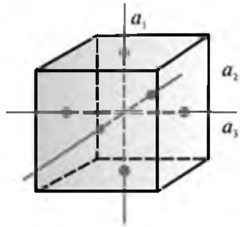


Рис. 286

Правильный тетраэдр — это тетраэдр, у которого все грани — правильные треугольники.

Примером правильного тетраэдра служит тетраэдр, образованный попарно не смежными вершинами куба (рис. 287). Более того, всякий правильный тетраэдр можно, таким образом, вписать в подходящий куб. Перемещения, при которых тетраэдр сохраняет свое положение, составляют часть тех перемещений, при которых сохраняет свое положение куб. Всего существует 24 таких перемещения, среди них — три поворота на 90°, четыре — на 120°, четыре — на 240°, шесть симметрий относительно плоскости (рис. 288). Эти перемещения позволяют отождествить все трехгранные углы правильного тетраэдра.

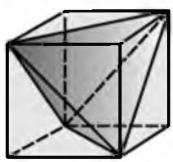


Рис. 287

Рис. 288

Следующий правильный многогранник — это октаэдр. Его вершинами служат центры граней куба; октаэдр является объединением двух четырехугольных пирамид (рис. 289).

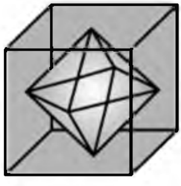


Рис. 289

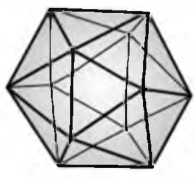


Рис. 290

Рис. 291

Перемещения, которые не меняют положения куба, сохраняют и положение октаэдра. Более того, ими исчерпываются все перемещения, при которых октаэдр сохраняет свое положение.

Существует еще два типа многогранников, у которых перемещениями, сохраняющими многогранник в целом, можно поменять местами любые две вершины, любые два ребра и любые две грани. Это икосаэдр (рис. 290) и додекаэдр.

Икосаэдр имеет 12 вершин, 30 ребер и 20 граней (являющихся правильными треугольниками). Додекаэдр имеет 20 вершин, 30 ребер и 12 граней (являющихся правильными пятиугольниками; рис. 291). Эти два типа многогранников тесно связаны между собой: центры граней икосаэдра служат вершинами додекаэдра, а центры граней додекаэдра служат вершинами икосаэдра. Икосаэдр и додекаэдр обладают 120 перемещениями (среди них 60 поворотов), сохраняющими положение этих тел.

Итак, мы ознакомились с пятью многогранниками, которые мы назвали правильными. Это правильный тетраэдр, куб, октаэдр, икосаэдр и додекаэдр. Можно по-разному определить «правильность» многогранника, например, заметив, что их грани правильные многоугольники. Однако существует много других многогранников, грани которых являются правильными многоугольниками. Некоторые из них изображены на рис. 292. Обычно определяют правильность многогранника так.

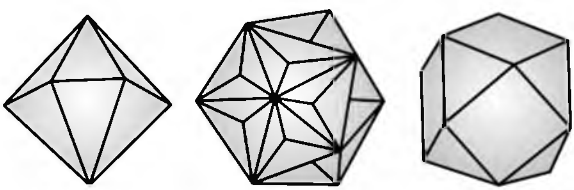


Рис. 292

Выпуклый многогранник называется правильным, если его грани — равные правильные многоугольники и в каждой его вершине сходится одинаковое число граней.

Оказывается, что этому определению «правильности» уже никакие другие многогранники, кроме указанных выше пяти, не удовлетворяют. Это можно строго доказать с помощью теоремы Эйлера.

Доказательство. Пусть в каждой вершине правильного многогранника сходятся m ребер и каждая грань — многоугольник с n ребрами. Пусть у многогранника всего *В* вершин, *Р* ребер и *Г* граней. Подсчитаем разными способами число ребер. Всего есть *Г* граней, в каждой *n* ребер, всего *Гn* ребер, но каждое сосчитано два раза, так как оно принадлежит двум граням. Итак, $Гn = 2P$.

С другой стороны, всего есть *В* вершин, в каждой сходится *m* ребер, но снова каждое ребро засчитано дважды, так как оно соединяет две вершины. Итак, $Bm = 2P$.

Подставим в соотношение Эйлера $Г + B = P + 2$ вместо *Г* и *В* их выражения через *m*, *n* и *P*:

$$\frac{2P}{m}+\frac{2P}{n}=P+2.$$

Нам достаточно даже неравенства $\frac{2P}{m}+\frac{2P}{n}>P$. Сокращая его на $2P$, получаем $\frac{1}{m}+\frac{1}{n}>\frac{1}{2}$. Числа *m* и *n* больше или равны трем. Оба они не могут быть больше трех, так как уже $\frac{1}{4}+\frac{1}{4}=\frac{1}{2}$, и при $m, n \geq 4$ получим: $\frac{1}{m}+\frac{1}{n}\leq\frac{1}{2}$. Есть возможности $m_1 = n_1 = 3$; $m_2 = 3$, $n_2 = 4$; $m_3 = 4$, $n_3 = 3$; $m_4 = 5$, $n_4 = 3$; $m_5 = 3$, $n_5 = 5$. Если же одно из чисел равно трем, а другое больше или равно шести, то

$$\frac{1}{m}+\frac{1}{n}\leq\frac{1}{3}+\frac{1}{6}=\frac{1}{2}.$$

Итак, у неравенства только пять решений. Они соответствуют известным нам пяти типам правильных многогранников.

Многогранные углы и многогранники

В этом параграфе мы уже не раз рассматривали многогранные углы: трехгранные углы — это углы, образованные сходящимися в одной вершине гранями тетраэдра, параллелепипеда, додекаэдра; четырехгранные углы — это углы, образованные сходящимися в одной вершине гранями октаэдра; пятигранные углы — это углы, образованные сходящимися в одной вершине гранями додекаэдра; наконец, угол при вершине *n*-угольной пирамиды (образованной гранями пирамиды) — это пример *n*-гранного угла. Последний пример дает нам *n*-гранные углы с любым $n \geq 3$.

Вспоминая определение пирамиды, мы приходим к следующей конструкции многогранного угла: Пусть даны n-угольник и точка O, не лежащая в плоскости n-угольника; тогда объединение лучей, соединяющих точку O с точками n-угольника, является n-гранным углом. Эта конструкция (называется *конической*) еще не исчерпывает всего запаса многогранных углов — она не охватывает двугранные углы и некоторые n-гранные углы с $n > 2$, однако объединения углов, получающихся с помощью этой конструкции и имеющих общую вершину, уже дают все возможные многогранные углы (на рис. 293 изображен многогранный угол, который не получается конической конструкцией; на рисунке он представлен как объединение трех углов, строящихся с помощью конической конструкции).

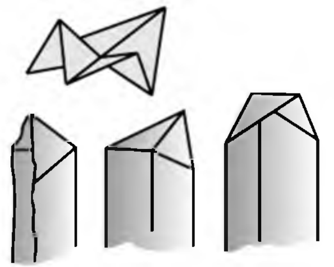


Рис. 293

Наиболее простой тип многогранника — это *выпуклые многогранники*. Пусть задано несколько плоскостей в пространстве; они разбивают пространство на части, и те из частей, которые ограничены, называются выпуклыми многогранниками (например, тетраэдр, параллелепипед, правильные многогранники). Чтобы убедиться в их выпуклости, достаточно провести плоскости, в которых лежат грани этих многогранников. Часто наиболее удобным способом проверки выпуклости служит следующий критерий: многогранник выпуклый тогда и только тогда, когда он вместе с любыми своими двумя точками содержит целиком весь отрезок, соединяющий эти точки. С помощью этого критерия легко убедиться в том, что призма, изображенная на рис. 294, не выпуклая.

Итак, выпуклые многогранники не исчерпывают всего запаса многогранников. Однако любой многогранник можно сложить, как из кирпичей, из выпуклых многогранников (рис. 295).

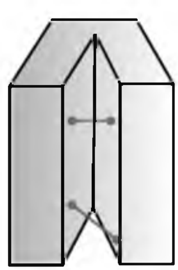


Рис. 294

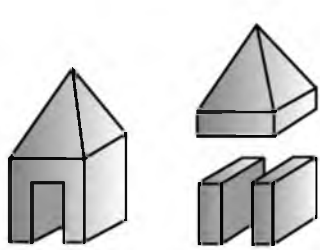


Рис. 295

Контрольные вопросы и задания

1. Что такое многогранник?
2. В чем состоит теорема Эйлера о выпуклых многогранниках?
3. Сколько есть различных типов правильных многогранников?
4. Какие правильные многоугольники могут быть гранями правильных многогранников?
5. Какой многогранник называется выпуклым?

ГЛАВА 7

МАТЕМАТИЧЕСКИЙ АНАЛИЗ

§ 29. Последовательности и их пределы

Задание числовой последовательности

Последовательность — это занумерованный ряд объектов. Можно строить последовательности чисел, функций, векторов, рассматривать последовательности событий, утверждений и др.

В качестве номеров, меток, определяющих порядок членов последовательности, обычно берут натуральные числа 1, 2, 3, ... Эти числа сами задают простейшую последовательность, порядок членов которой мы считаем известным.

Занумерованный ряд чисел a_1, a_2, a_3, ..., an, ... называется числовой последовательностью.

Наиболее простой способ задания последовательности — это ее задание с помощью формулы общего члена, т.е. формулы, явно выражающей зависимость n-го члена последовательности от n.

Например, формула $a_n = 2n$ задает последовательность четных чисел 2, 4, 6, 8, ...

Другим важным способом задания последовательности является так называемый *рекуррентный* способ, при котором задается выражение, связывающее n-й член последовательности с одним или несколькими предыдущими. Слово *рекуррентный* происходит от латинского слова *рекурсия*, что означает возврат. Вычисляя новый, очередной член последовательности, мы как бы возвращаемся назад, к уже вычисленным, предыдущим членам.

Примеры

1. Рекуррентное соотношение $a_n = a_{n-1} + 2$ вместе с условием $a_1 = 1$ задает арифметическую прогрессию с первым членом 1 и разно-

стью 2: 1, 3, 5, 7, ... Это не что иное как последовательность нечетных чисел.

2. Рекуррентное соотношение $a_n = 2a_{n-1}$ вместе с условием $a_1 = 1$ задает геометрическую прогрессию с первым членом 1 и знаменателем 2: 1, 2, 2^2, 2^3, ... Это не что иное как последовательность степеней двойки, начиная с нулевой степени.

   Кстати, иногда члены последовательности удобно нумеровать с нуля или вообще выбирать другой способ нумерации.

3. Рекуррентное соотношение $a_n = a_{n-1} + a_{n-2}$ вместе с условием $a_0 = 0$, $a_1 = 1$ задает последовательность чисел Фибоначчи: 0, 1, 1, 2, 3, 5, 8, 13, 21, ...

Последовательность может быть задана словесным описанием, в котором определяется процесс построения членов последовательности.

Например, описание «пусть a_n — это n-е простое число» задает последовательность 2, 3, 5, 7, 11, 13, ..., члены которой берутся из таблицы простых чисел или вычисляются каким-либо другим способом (например, с помощью решета Эратосфена).

Последовательность является дискретным вариантом понятия функции. В отличие от привычной функции типа $y = f(x)$, аргумент которой x определен на некотором числовом промежутке, последовательность $a_1, a_2, ..., a_n, ...$ можно считать функцией, аргумент которой n принимает дискретный ряд значений $n = 1, 2, 3, ...$ Часто ее n-й член можно выразить как значение некоторой обычной функции $y = f(x)$ для $x = n$: $a_n = f(n)$.

Примеры

1. 1, 4, 9, 16, ... — последовательность квадратов чисел натурального ряда. Формула общего члена: $a_n = n^2$. Можно считать, что a_n — это значение функции $y = x^2$ при $x = n$.

2. 1, $\frac{1}{2}$, $\frac{1}{3}$, ... — последовательность обратных чисел. Формула общего члена: $a_n = \frac{1}{n}$. Аналогично предыдущему примеру здесь $a_n = f(n)$, где $f(x) = \frac{1}{x}$.

3. 1, 2, 6, 24, ... — последовательность факториалов. Формула общего члена: $a_n = n!$.

Нам неизвестна функция $y = f(x)$, заданная при всех $x \geq 0$, значения которой при $x = n$ равнялись бы $n!$. Однако такая функция существует. Ее построил Л. Эйлер с помощью интегралов.

4. Последовательность десятичных приближений к числу $\sqrt{2}$. Зная способ вычисления десятичных знаков числа $\sqrt{2} = 1{,}414...$, мы получим последовательность 1; 1,4; 1,41; 1,414; ...
5. Последовательность цифр в десятичной записи числа π.

Первые 707 знаков этого числа были вычислены еще в XVIII в. (правда, в вычисления закралась ошибка). Сейчас компьютер легко выдает достаточно большое число знаков π:

$$3, 1, 4, 1, 5, 9, 2, 6, 5, 3, 6, ...$$

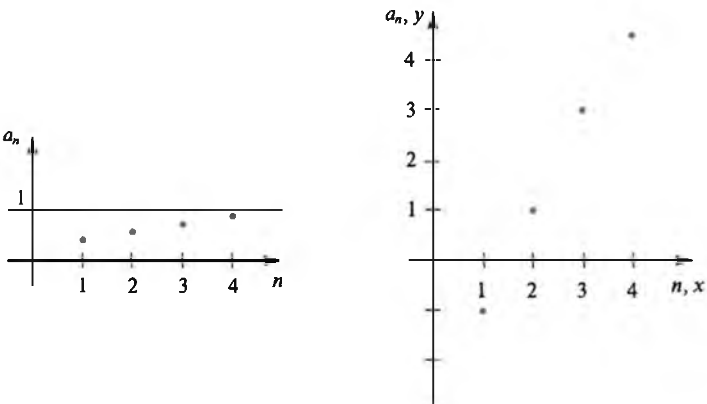
Числовые последовательности могут обладать свойствами, которые мы обсуждали при изучении обычных функций.

Числовая последовательность называется *возрастающей*, если каждый ее член больше предыдущего, иными словами, если для всякого $n > 1$ верно неравенство $a_n > a_{n-1}$.

Аналогично дается определение *убывающей* числовой последовательности.

Вместе возрастающие и убывающие последовательности называются *монотонными* последовательностями.

Последовательность a_1, a_2, ... можно изобразить «графиком», который будет состоять из отдельных точек координатной плоскости (рис. 296). Так же, как и для обычных функций, по графику можно судить о различных свойствах последовательностей. Возрастающие и убывающие последовательности изображаются точками, лежащими на графиках монотонных функций.



а *б*

Рис. 296

Ограниченные последовательности

Последовательность a_1, a_2, a_3, ... называется *ограниченной*, если для ее членов можно указать общую *границу*, т.е. такое число C, что неравенство $|an| \leq C$ выполняется для всех номеров n.

Если последовательность является возрастающей, то для ее ограниченности достаточно найти число C такое, что $a_n \leq C$ при всех n. Наоборот, для ограниченности убывающей последовательности достаточно проверить неравенство вида $a_n \geq C$, которое должно выполняться для всех n. Вообще, если для всех членов последовательности выполняется неравенство $a_n \leq C$ ($a_n \geq C$), то говорят, что она ограничена сверху (снизу). Если мы говорим об ограниченной последовательности, то ясно, что она ограничена как сверху, так и снизу.

Действия над последовательностями

Так же, как над произвольными функциями (заданными на одном и том же множестве), над последовательностями можно производить арифметические операции: сложение (вычитание) и умножение (деление).

Суммой двух последовательностей a_1, a_2, a_3, ... и b_1, b_2, b_3, ... называется последовательность c_1, c_2, c_3, ..., образованная суммами соответствующих членов: $c_1 = a_1 + b_1$, $c_2 = a_2 + b_2$, $c_3 = a_3 + b_3$, ...

Аналогично перемножаются две последовательности: $d_1 = a_1 \cdot b_1$, $d_2 = a_2 \cdot b_2$, $d_3 = a_3 \cdot b_3$, ...

Если последовательность b_1, b_2, ... *постоянна*, т.е. если $b_n = b$ для любого n, то произведение последовательностей a_1, a_2, ... и b_1, b_2, ... выглядит так: ba_1, ba_2, ... и называется произведением постоянного числа b на последовательность a_1, a_2, ...

Примеры

1. $1, 4, 9, 16, ..., a_n = n^2$.

   Эта последовательность является возрастающей (аналогично тому, что функция $y = x^2$ возрастает при $x \geq 0$). Она не является ограниченной, так как n^2 может стать сколь угодно большим.

2. $1, \frac{1}{2}, \frac{1}{3}, ..., a_n = \frac{1}{n}$.

   Эта последовательность является убывающей: $1 > \frac{1}{2} > \frac{1}{3} > ...$, что аналогично убыванию функции $y = \frac{1}{x}$ для $x > 0$. Последовательность $a_n = \frac{1}{n}$ является ограниченной: $|a_n| \leq 1$. Разумеется, так как эта последовательность убывает, то каждый ее член меньше первого: $a_n \le a_1 = 1$. Важно то, что она ограничена снизу: $a_n > 0$.

3. $a_n = n!$
4. $a_n = 2n - 1$.
5. $a_n = 2^{n-1}$.

   Все эти три последовательности возрастают и не являются ограниченными.
6. a_n — n-е десятичное приближение с недостатком к числу $\sqrt{2}$. Эта последовательность возрастает и ограничена: $a_n < \sqrt{2}$.
7. Последовательность цифр в десятичной записи, конечно, не может быть монотонной. Она всегда ограничена, так как $0 \le a_n \le 9$.

С каждой последовательностью $a_1, a_2, a_3, \ldots$ можно связать две новых последовательности.

Последовательность сумм:
$s_1 = a_1$;
$s_2 = a_1 + a_2$;
$s_3 = a_1 + a_2 + a_3$ и т.д.
Последовательность сумм можно определить рекуррентно:

$$s_1 = a_1;\ s_n = s_{n-1} + a_n.$$

Последовательность разностей:
$c_1 = a_2 - a_1$;
$c_2 = a_3 - a_2$;
$c_3 = a_4 - a_3$ и т.д.
Построим последовательности разностей для нескольких примеров.

1. (a_n): 1, 2, 3, 4, ... $\quad a_n = n$;
   (c_n): 1, 1, 1, ... $\quad c_n = 1$.
2. (a_n): 1, 3, 5, 7, ... $\quad a_n = 2n - 1$;
   (c_n): 2, 2, 2, ... $\quad c_n = 2$.
3. (a_n): $1, 2, 2^2, 2^3, \ldots$ $\quad a_n = 2^{n-1}$;
   (c_n): 1, 2, 4, ... $\quad c_n = 2^{n-1}$.
4. a_n: $1, 2^2, 3^2, 4^2, \ldots$ $\quad a_n = n^2$;
   c_n: 3, 5, 7, ... $\quad c_n = 2n + 1$.
5. a_n: 1, 1, 2, 3, 5, 8, ... $\quad a_n$ — числа Фибоначчи;
   c_n: 0, 1, 1, 2, 3, ... $\quad c_n$ — те же числа со сдвинутым номером.
6. a_n: $1, 2^3, 3^3, \ldots$ $\quad a_n = n^3$;
   c_n: 7, 19, 37, ... $\quad c_n = (n + 1)^3 - n^3 = 3n^2 + 3n + 1$.

Рассматривая эти примеры, можно заметить несколько закономерностей. Если общий член последовательности записывается многочленом от n, то степень общего члена разностей будет на единицу меньшей. В примерах 1 и 2 член a_n линейно зависит от n, в примере 4 квадратично, в примере 6 — зависимость кубическая. Соответствующие последовательности разностей постоянны (степень 0), линейны или квадратичны. В последовательности 3 общий член задан показательной функцией. Общий член последовательностей разностей имеет тот же вид.

Аналогичные наблюдения можно сделать и для последовательности чисел Фибоначчи. Оказывается, что имеется общий закон — если последовательность задается как показательная функция от n, то последовательность разностей будет пропорциональна той же показательной функции.

Это наводит на мысль о том, что построение последовательности разностей и ее свойства являются дискретным аналогом вычисления производной. Аналогично суммирование последовательности аналогично другой операции математического анализа — интегрированию.

Математическая индукция

Различные утверждения (теоремы) о последовательностях часто доказывают рассуждением, которое называется *математической индукцией*.

Пусть нам надо доказать утверждение вида: для каждого натурального числа n верно утверждение $P(n)$.

Метод математической индукции, применяемый для его доказательства, состоит в следующем.

1. Утверждение проверяется для начального значения n, т.е. доказывается утверждение $P(1)$.

2. Доказывается условное утверждение: если верно $P(n)$, то верно $P(n+1) : P(n) \Rightarrow P(n+1)$.

Первый шаг в этом доказательстве называется *базой индукции*, второй шаг называется *индукционным переходом*.

Применим метод математической индукции в простом случае для доказательства формулы для суммы квадратов первых n натуральных чисел:

$$s_n = 1^2 + 2^2 + \ldots + n^2 = \frac{n(n+1)(2n+1)}{6}.$$

Прежде всего, проверяем справедливость этой формулы для первых значений n:

$n = 1$ $\qquad s_1 = 1 = \frac{1 \cdot 2 \cdot 3}{6}$;

$n = 2$ $\qquad s_2 = 1 + 4 = 5 = \frac{2 \cdot 3 \cdot 5}{6}$.

Затем предполагаем, что формула справедлива для суммы $n - 1$ члена и доказываем, что она будет справедлива и для суммы n членов. Тогда от начальных значений $n = 1$, 2 мы сможем последовательно переходить дальше: $2 \Rightarrow 2 + 1 = 3$, $3 \Rightarrow 3 + 1 = 4$ и т.д., т.е. установим справедливость формулы для любого натурального n. Осталось провести выкладки.

Дано: $s_{n-1} = \frac{(n-1)(n-1+1)\left(2(n-1)+1\right)}{6} = \frac{(n-1) \cdot n \cdot (2n+1)}{6}$.

Доказать: $s_n = \frac{n(n+1)(2n+1)}{6}$.

Доказательство: $s_n = s_{n-1} + a_n = \frac{(n-1) \cdot n \cdot (2n-1)}{6} + n^2 = \frac{n}{6}((n-1)$ $(2n-1) + 6n) = \frac{n}{6}(2n^2 + 3n + 1) = \frac{n(n+1)(2n+1)}{6}$, что и требовалось доказать.

Важность каждого из двух пунктов при доказательстве теорем методом математической индукции поясним шуточным примером.

Допустим, что вы должны доехать на автобусе до некоторого места, находящегося от вас за несколько остановок. Вам нужно:

1) сесть в автобус;
2) знать, что если автобус приехал на какую-то остановку, то он приедет и на следующую.

Второе условие гарантирует, что если вы начали движение, то оно вас обязательно приведет к цели. Если же оно соблюдается, но вы не сели в автобус, то, разумеется, вы никуда не приедете.

Примеры

Чаще всего метод математической индукции применяется для доказательства формул, содержащих переменное натуральное число, скажем, n. Мы начали с доказательства формулы для суммы квадратов натуральных чисел. Докажем аналогичным рассуждением формулу суммы кубов первых n натуральных чисел:

$$s_n = 1^3 + 2^3 + \ldots + n^3 = \frac{n^2(n+1)^2}{4};$$

а) проверяем (устно) справедливость формулы при $n = 1$.

б) проводим индукционный переход. Для этого надо суметь записать формулу для двух последовательных значений натурального аргумента, например, n и $n + 1$ или $k - 1$ и k. Выражение для s_n нам дано. Подставим вместо n число $n + 1$, получим $s_{n+1} = \dfrac{(n+1)^2(n+2)^2}{4}$.

Итак, докажем, что $s_n = \dfrac{n^2(n+1)^2}{4} \Rightarrow s_{n+1} = \dfrac{(n+1)^2(n+2)^2}{4}$.

Переход основан на знании содержательной связи между s_n и s_{n+1}: $s_{n+1} = s_n + (n + 1)^3$. Подставляя в эту формулу условие теоремы $S_n = \dfrac{n^2(n+1)^2}{4}$, преобразуем ее к нужному виду: $\dfrac{n^2(n+1)^2}{4} + (n+1)^3 = \dfrac{(n+1)^2}{4}(n^2 + 4n + 4) = \dfrac{(n+1)^2(n+2)^2}{4}$, что и требовалось доказать.

Индуктивные рассуждения используют не только в алгебре. Приведем пример.

Доказать, что n прямых на плоскости в общем положении (т.е. никакие две из них не параллельны и никакие три не пересекаются в одной точке) делят плоскость на $a_n = \dfrac{n^2 + n + 2}{2}$ частей.

Доказательство. Начнем с базы индукции. Пусть $n = 1$. Одна прямая делит плоскость на две части. Проверяем формулу: $a_1 = \dfrac{1+1+2}{2} = 2$. Аналогично две (не параллельные) прямые делят плоскость на четыре части, и действительно, $a_2 = \dfrac{2^2 + 2 + 2}{2} = 4$. Аналогично $a_3 = \dfrac{3^2 + 3 + 2}{2} = 7$, что легко проверяется построением.

Выполняем индукционный переход. *Предположим*, что n прямых в общем положении делят плоскость на $\dfrac{n^2 + n + 2}{2}$ частей. *Докажем*, что $n + 1$ прямые делят плоскость на $\dfrac{(n+1)^2 + (n+1) + 2}{2}$ частей. Посмотрим геометрически, сколько новых областей добавляется при проведении $(n + 1)$-й прямой. Начнем проводить $(n + 1)$-ю прямую «издалека, из бесконечности». Когда она «встретит» первую из проведенных ранее n прямых, добавится одна новая область (рис. 297). Затем при подходе к каждой новой прямой добавится одна новая область. Таких «подходов» будет n, так как по условию мы пересечем *каждую* из n прямых (параллельных прямых не может быть) и каждую из них отдельно от другой (общих точек пересечения не может быть). После пересечения последней из n прямых, мы добавим в конце еще одну, $(n + 1)$-ю новую область.

Итак, число областей при проведении $(n + 1)$-й прямой увеличилось на $n + 1$. Осталось сделать выкладки. Зная, что n прямых разделили плоскость на $\frac{n^2+n+2}{2}$ области, мы получаем, что $n + 1$ прямая разделит плоскость на $\frac{n^2+n+2}{2} + (n + 1)$ область. Но $\frac{n^2+n+2}{2} + (n + 1) = \frac{(n^2+2n+1)+(n+1)+2}{2} = \frac{(n+1)^2+(n+1)+2}{2}$, что и требовалось доказать.

Термины индукция и дедукция часто противопоставляют друг другу, понимая под индуктивным умозаключением вывод, следующий из рассмотрения достаточно большого числа частных случаев, а под дедуктивным умозаключением вывод, следующий из исходных посылок с помощью логических правил вывода. В этом смысле метод математической индукции является **дедуктивным** методом, коль скоро он позволяет провести индукционный переход. Поэтому его иногда называют методом полной индукции, подчеркивая полноту доказанных утверждений вида $P(n)$ при всех значениях n.

Метод математической индукции очень полезен в тех случаях, когда ответ на вопрос, содержащий натуральный параметр n, неизвестен заранее. Тогда приходится угадывать ответ, строя правдоподобные гипотезы на основе анализа частных случаев (базы индукции).

Например, пытаясь найти формулу для суммы первых n нечетных чисел, т.е. найти $s_n = 1 + 3 + ... + 2n - 1$, легко заметить закономерность, вычисляя первые значения s_n:

$$s_1 = 1; s_2 = 4; s_3 = 9; s_4 = 16; s_5 = 25, ...$$

Напрашивается ответ: $s_n = n^2$. Проверка его с помощью индукционного перехода уже не представляет труда: $s_n = n^2 \Rightarrow s_{n+1} = (n + 1)^2$. Действительно, $s_{n+1} = s_n + 2n + 1 = n^2 + 2n + 1 = (n + 1)^2$, что и требовалось доказать.

Пример

Дана последовательность чисел Фибоначчи 1, 1, 2, ...; $a_{n+2} = a_{n+1} + a_n$. Выразить сумму s_n ее первых n членов.

Для решения сначала выпишем достаточно много чисел Фибоначчи.

$$1, 1, 2, 3, 5, 8, 13, 21, 34, 55, 89, 144, 233, 377, ...$$

Теперь выпишем последовательности частичных сумм.

$$1, 2, 4, 7, 12, 20, 33, 54, 88, 143, 232, ...$$

Заметим, что это последовательность тех же чисел Фибоначчи, уменьшенных на 1, и сдвинутая на два номера.
Строим гипотезу $s_n = a_1 + a_2 + ... + a_n = a_{n+2} - 1$.
Наши вычисления дают подтверждение сделанной гипотезы для первых двух номеров. Проведем индукционный переход.
Предположим, что $s_n = a_{n+2} - 1$. *Докажем*, что $s_{n+1} = a_{n+3} - 1$.
По построению сумм $s_{n+1} = s_n + a_{n+1} = a_{n+2} - 1 + a_{n+1} = a_{n+1} + a_{n+2} - 1 = a_{n+3} - 1$, что и требовалось доказать.

Разумеется, не всегда дело обстоит так просто — бывает так, что и правдоподобный ответ подобрать трудно, и переход провести непросто.

Предел последовательности

Двадцать пять веков (!) люди размышляют об одном из рассуждений древнегреческого философа Зенона, жившего в V в. до н.э. Приведем цитату из «Войны и мира» Л.Н. Толстого: «Известен так называемый софизм древних, состоящий в том, что Ахиллес никогда не догонит впереди идущую черепаху, несмотря на то, что Ахиллес идет в десять раз скорее черепахи: как только Ахиллес пройдет пространство, отделяющее его от черепахи, черепаха пройдет впереди его одну десятую этого пространства; Ахиллес пройдет эту десятую, черепаха пройдет одну сотую и так до бесконечности. Задача эта представлялась древним неразрешимой».

Нам эта задача кажется до смешного простой: Ахиллес пробегает отрезки, равные 1, $\frac{1}{10}$, $\frac{1}{10^2}$, ... от начального расстояния. Сложив бесконечно убывающую прогрессию $1+\frac{1}{10}+\frac{1}{100}+...=\frac{1}{1-0{,}1}=\frac{10}{9}$, мы вычислим, какой путь пробегает Ахиллес до встречи с черепахой.

Однако в этом решении далеко не все так просто. Это решение основано на некотором бесконечном процессе, а представление о бесконечности содержит в себе много противоречий. Анализ этих противоречий — задача скорее философии, чем математики.

Для того чтобы обосновать рассуждения, связанные с бесконечными процессами, математики XIX в. создали «теорию пределов», которая позволила поставить на прочный фундамент многие результаты предыдущих веков. Такие математические понятия как сумма бесконечного ряда, производная, интеграл, непрерывность могут быть определены с помощью понятия предела, что позволит строго доказывать и применять свойства этих понятий.

Одним из наиболее простых предельных переходов является вычисление суммы бесконечно убывающей геометрической прогрессии. Он основан на рассмотрении частичных сумм $s_n = a_1(1 + q + \ldots + q^{n-1}) = a_1 \dfrac{1-q^n}{1-q}$. Центральным местом рассуждения является утверждение о том, что «предел последовательности q^n при $|q| < 1$ равен нулю». Попробуем уточнить понятие предела последовательности.

Число A называется пределом последовательности $a_1, a_2, \ldots$, если, начиная с некоторого места, все члены этой последовательности будут сколь угодно мало отличаться от A. Обозначение: $A = \lim\limits_{n\to\infty} a_n$.

Вычислим пределы некоторых последовательностей.

1. $a_n = \dfrac{1}{n}$. Ясно, что пределом этой последовательности будет число 0. Действительно, взяв произвольное число $\alpha > 0$, мы можем найти такой номер последовательности, после которого каждый член a_n будет меньше α (т.е. $|a_n - 0| = \dfrac{1}{n} < \alpha$). Действительно, $\dfrac{1}{n} < \alpha \Leftrightarrow n > \dfrac{1}{\alpha}$, так что достаточно взять любое n, большее числа $\dfrac{1}{\alpha}$. Итак, $\lim\limits_{n\to\infty} \dfrac{1}{n} = 0$.

2. $\lim\limits_{n\to\infty} \dfrac{1}{n^k} = 0$ при любом $k > 0$. Действительно, мы можем «решить неравенство» $\dfrac{1}{n^k} < \alpha \Leftrightarrow n^k > \dfrac{1}{\alpha} \Leftrightarrow n > \dfrac{1}{\alpha^{\frac{1}{k}}}$. Справа стоит вполне определенное положительное число, которое указывает нам, с какого места число $\dfrac{1}{n^k}$ станет меньше наперед заданного числа $\alpha > 0$.

3. $\lim\limits_{n\to\infty} q^n = 0$ при $|q| < 1$.

Сейчас уже «решить неравенство» $|q|^n < \alpha$ непросто — для этого нужны логарифмы. (Запишем это решение: $|q|^n < \alpha \Leftrightarrow n \lg|q| < \lg\alpha \Leftrightarrow$ $\Leftrightarrow n > \dfrac{\lg\alpha}{\lg|q|}$. Знак неравенства поменяется, потому что $\lg|q| < 0$ при $|q| < 1$. Отрицательным будет и $\lg\alpha$, если $\alpha < 1$, и мы получим положительную границу для n.) Однако нам не обязательно находить самое первое значение номера n, начиная с которого число $|q|^n$ станет меньше данного числа α. Достаточно дать *оценку* для степени $|q|^n$. Это можно сделать, например, с помощью неравенства Бернулли: $(1 + a)^n > 1 + na$, верного при любом $a > 0$. Возьмем $1+a=\frac{1}{|q|} \Leftrightarrow a=\frac{1-|q|}{|q|}>0$. Получим $\frac{1}{|q|^n}>1+na \Leftrightarrow |q|^n<\frac{1}{1+na}<\frac{1}{an}$. Ясно, что дробь $\frac{1}{an}$ $(a > 0)$ может быть сделана сколь угодно малой.

Вычислению пределов последовательностей помогают следующие простые свойства.

1. $\lim\limits_{n\to\infty}(a_n+b_n)=\lim\limits_{n\to\infty}a_n+\lim\limits_{n\to\infty}b$.
2. $\lim\limits_{n\to\infty}ca_n=c\lim\limits_{n\to\infty}a_n$.
3. $\lim\limits_{n\to\infty}a_n\cdot b_n=\lim\limits_{n\to\infty}a_n\cdot\lim\limits_{n\to\infty}b_n$.
4. $\lim\limits_{n\to\infty}\frac{a_n}{b_n}=\frac{\lim\limits_{n\to\infty}a_n}{\lim\limits_{n\to\infty}b_n}$, если $\lim\limits_{n\to\infty}b_n\neq 0$.

Во всех этих равенствах предполагается, что все написанные пределы существуют, или, как говорят математики, все последовательности являются *сходящимися*.

Примеры

Вычислим несколько выражений с помощью сформулированных свойств.

1. $\lim\limits_{n\to\infty}\frac{n}{n+1}$.

Можно поступить так: $\frac{n}{n+1}=\frac{n+1-1}{n+1}=1-\frac{1}{n+1}$.

Будем считать число 1 общим членом «постоянной» последовательности, предел которой, конечно, равен 1:

$$\lim_{n\to\infty}\frac{n}{n+1}=\lim_{n\to\infty}1-\frac{1}{n+1}=\lim_{n\to\infty}1-\lim_{n\to\infty}\frac{1}{n+1}=1-0=1.$$

Итак, $\lim\limits_{n\to\infty}\frac{n}{n+1}=1$, что мы, разумеется, знали и раньше (например, когда проводили горизонтальную асимптоту для построения графика функции $y=\frac{x}{x+1}$).

Можно было поступить по-другому: $\dfrac{n}{n+1}=\dfrac{1}{1+\dfrac{1}{n}}$ и теперь воспользоваться свойством пределов:

$$\lim_{n\to\infty}\frac{n}{n+1}=\lim_{n\to\infty}\frac{1}{1+\dfrac{1}{n}}=\frac{\lim\limits_{n\to\infty}1}{\lim\limits_{n\to\infty}1+\dfrac{1}{n}}=\frac{1}{1+0}=1.$$

2. $$\lim_{n\to\infty}\frac{2n^2-3n+1}{3n^2+n+2^n}=\lim_{n\to\infty}\frac{n^2\left(2-\dfrac{3}{n}+\dfrac{1}{n^2}\right)}{n^2\left(3+\dfrac{1}{n}+\dfrac{2}{n^2}\right)}=\lim_{n\to\infty}\frac{2-\dfrac{3}{n}+\dfrac{1}{n^2}}{3+\dfrac{1}{n}+\dfrac{2}{n^2}}=\frac{2-0+0}{3+0+0}=\frac{2}{3}.$$

Существование предела

Рассмотрим такое число $A=\sqrt{2+\sqrt{2+\sqrt{2+\ldots}}}$. Что такая запись может обозначать? Можно начать с числа $a_1=\sqrt{2}$, затем перейти к числу $\sqrt{2+\sqrt{2}}$, затем под последним корнем прибавить $\sqrt{2}$ и так «продолжать до бесконечности». Мы столкнулись с последовательностью a1, a2, ..., у которой $a_1=\sqrt{2}$, а каждый следующий вычисляется по рекуррентной формуле $a_{n+1}=\sqrt{2+a_n}$ или $a_{n+1}^2=a_n+2$. Интересующее нас число надо понимать как предел последовательности (ап), т.е. $A=\sqrt{2+\sqrt{2+\sqrt{2+\ldots}}}=\lim\limits_{n\to\infty}a_n$.

Чему же равно число A и существует ли оно? «Перейдем в равенстве $a_{n+1}^2=a_n+2$ к пределу» (и чуть позже разберемся в смысле этих слов). Получим $A^2=A+2$, или $A^2-A-2=0$, откуда $A=2$ или $A=-1$. Так как, судя по записи с радикалами, $A>0$, то выберем значение $A=2$, т.е. $\sqrt{2+\sqrt{2+\sqrt{2+\ldots}}}=2$.

Начнем с переходов к пределу. Они основаны на упоминавшихся ранее правилах перехода к пределу при арифметических действиях с последовательностями.

Равенство $a_{n+1}^2=a_n+2$ можно понимать как арифметическую связь между последовательностями: последовательность $b_n=a_{n+1}$ — это та же последовательность (a_n), только начинающаяся со второго ее члена, число 2 можно понимать как «постоянную последовательность» c_n, где $c_n=2$ при всех n. Теперь рекуррентное соотношение запишем как

$b_n \cdot b_n = a_n + c_n$. Применяя правило перехода к пределу, получим $A^2 = A + 2$. При этом использованы два очевидных соображения: пределы последовательностей a_n и $b_n = a_{n+1}$ равны между собой, а предел постоянной последовательности равен значению каждого ее члена.

Казалось бы, все обосновано, и задача решена. Применим метод решения к последовательности, заданной рекуррентным соотношением $a_{n+1} = 2a_n - 1$. Обозначим $\lim\limits_{n\to\infty} a_n$ через A и перейдем к пределу: $A = 2A - 1 \Rightarrow A = 1$. Действительно, если возьмем $a_1 = 1$, то получим, что $a_2 = 2 \cdot 1 - 1 = 1$, $a_3 = 1$ и т.д., т.е. $\lim\limits_{n\to\infty} a_n = 1$. Однако, если мы начнем с $a_1 = 2$, то получим последовательность 2, 3, 5, 9, 17, ..., общий член которой легко угадывается $a_n = n^2 + 1$ (и легко проверяется по индукции), но никакого предела у последовательности $a_n = n^2 + 1$ быть не может. Наша ошибка состоит в том, что сформулированные (без доказательства) правила обращения с пределами предполагают *существование* пределов всех существующих последовательностей.

Понятно, что одни последовательности имеют пределы (*сходящиеся последовательности*), другие — нет (*расходящиеся последовательности*). Для доказательства сходимости последовательности часто бывает полезен следующий признак.

Признак сходимости последовательности

Если последовательность монотонна и ограниченна, то она имеет предел.

Этот признак легко иллюстрируется с помощью числовой оси: двигаясь по ней в одну сторону и не имея возможности перейти через поставленный барьер, мы неограниченно приблизимся к некоторой точке числовой оси (рис. 298). Это утверждение (или ему эквивалентное) обычно принимается в качестве одной из аксиом множества вещественных чисел (аксиома непрерывности).

Рис. 298

В нашем примере с корнями последовательность (a_n) была возрастающей и ограниченной. Действительно, сначала проверим, что $a_n < 2$ при всех n. Применим индукцию: $a_1 = \sqrt{2} < 2$. Если $a_n < 2$, то $a_n + 2 < 4$ и $\sqrt{a_n + 2} < \sqrt{4} = 2$, что и утверждалось. Теперь докажем возрастание последовательности. Нам надо доказать, что $a_{n+1} = \sqrt{a_n + 2}$ больше, чем a_n, т.е. проверить неравенство $\sqrt{a_n + 2} > a_n$. Так как $a_n > 0$, то его можно возвести в квадрат и получить для проверки неравенство $a_n^2 - a_n - 2 < 0$. Решая неравенство $x^2 - x - 2 < 0$, получим промежуток $(-1; 2)$, в котором лежат числа последовательности $(0 < a_n < 2)$. Существование предела полностью доказано.

Важным геометрическим примером применения пределов последовательностей является вычисление длины окружности и площади круга как пределов периметров и площадей последовательностей многоугольников.

Пусть дан круг радиуса R. Рассмотрим последовательность M_n правильных n-угольников ($n \geq 3$), вписанных в эту окружность (рис. 299). Обозначим через p_n периметр M_n, а через S_n — его площадь. Легко представить себе (но непросто доказать), что последовательности p_n и S_n возрастают. Если p — длина окружности (границы взятого круга), а S — площадь круга, то $p_n < p$ и $S_n < S$. Это означает, что последовательности p_n и S_n монотонны и ограничены. Значит, по признаку сходимости, они должны иметь предел. Опять же «геометрически ясно», что $\lim\limits_{n\to\infty} p_n = p$ и $\lim\limits_{n\to\infty} S_n = S$, т.е. длину окружности и площадь круга можно вычислить как пределы последовательностей периметров и площадей правильных вписанных многоугольников.

Чтобы полностью разобраться в деталях приведенного рассуждения, надо проделать довольно большую работу. Наметим ее основные этапы.

1. Чтобы получить удобные формулы, связывающие два соседних многоугольника, берут последовательность не всех многоугольников, а только тех, которые получаются удвоением сторон. Например, взяв в качестве первого члена A_1 правильный треугольник $A_1 = M_3$, потом удваивают число сторон: $A_2 = M_{2\cdot 3}$, $A_3 = M_{2\cdot 3^2}$, и т.д. Так как каждый многоугольник является частью следующего (рис. 300), то теперь действительно очевидно возрастание последовательностей периметров и площадей.

Рис. 299

Рис. 300

2. Если понятия длины окружности p и площади круга S определены заранее, то неравенства $p_n < p$ и $S_n < S$ выполняются в силу свойств длин и площадей: длина объемлющей больше длины объемлемой для выпуклых линий, и площадь части фигуры меньше площади всей фигуры. Теперь надо доказать, что $\lim\limits_{n\to\infty} p_n = p$ и $\lim\limits_{n\to\infty} S_n = S$, т.е. что действительно последовательности периметров и площадей имеют своими пределами числа p и S. Для этого надо провести оценки разностей $p - p_n$ и $S - S_n$.

Заметим, что иногда рассмотренная нами конструкция используется для *определения* длины окружности и площади круга. В этом случае надо дать другую оценку для ограниченности p_n и S_n, использовав, например, многоугольники, описанные около данного круга.

3. Было бы интересно сравнить пределы последовательностей p_n и S_n с известными нам формулами длины окружности и площади круга. Для этого нам понадобятся приближенные значения синуса.

Пусть M_n — правильный n-угольник, вписанный в окружность радиуса R, a_n — его сторона, h_n — его апофема. Тогда $p_n = na_n$, $S_n = \frac{1}{2} na_nh_n$. Выразим a_n и h_n через R: $a_n = 2R\sin\frac{\pi}{n}$, $h_n = R\cos\frac{\pi}{n}$. Получим формулы $p_n = 2Rn\sin\frac{\pi}{n}$, $S_n = R^2n \sin \frac{\pi}{n} \cos \frac{\pi}{n} = = \frac{R^2}{2}n\sin\frac{2\pi}{n}$. При больших n (и, следовательно, малых углах $\frac{\pi}{n}$ и $\frac{2\pi}{n}$) имеем приближенные равенства $\sin \frac{\pi}{n} \approx \frac{\pi}{n}$, $\sin \frac{2\pi}{n} \approx \frac{2\pi}{n}$, что на языке пределов может быть записано так: $\lim\limits_{n\to\infty} n \cdot \sin \frac{\pi}{n} = \pi$ и $\lim\limits_{n\to\infty} n \cdot \sin \frac{2\pi}{n} = 2\pi$. Получаем известные формулы: $\lim\limits_{n\to\infty} p_n = 2\pi R$ и $\lim\limits_{n\to\infty} S_n = \pi R^2$.

Суммирование ряда

Возьмем кусок пирога. Отрежем от него половину, затем половину от оставшегося и т.д. Ясно, что на последовательных шагах мы отрезаем куски, равные $\frac{1}{2}, \frac{1}{4}, \frac{1}{8}$ и т.д. от всего пирога. Продолжая этот процесс «до бесконечности», мы, разумеется, исчерпаем весь пирог. Символически этот результат можно записать так:

$$\frac{1}{2}+\frac{1}{4}+\frac{1}{8}+\ldots=1.$$

Разумеется, на практике невозможно измельчать кусочек пирога до бесконечности, но математика давно придумала способ, как описывать подобные бесконечные процессы.

Рядом называется последовательность, у которой хотят найти сумму всех ее членов.

Запишем ряд: $a_1 + a_2 + a_3 + \ldots$

Вычислим n-ю частичную сумму s_n последовательности $a_1, a_2, a_3, \ldots$ Ее же называют частичной суммой ряда.

Пусть с ростом n частичная сумма sn неограниченно приближается к некоторому числу s. Тогда число s называют суммой ряда и записывают так:

$$a_1 + a_2 + a_3 + \ldots = s.$$

Можно сказать, что число s является пределом последовательности $s_1, s_2, \ldots$ частичных сумм: $s = \lim\limits_{n\to\infty} s_n$. Рассмотрим лишь несколько простых примеров.

Примеры

1. Исторически первый неочевидный (как говорят математики, нетривиальный) пример суммирования бесконечного числа слагаемых принадлежит Архимеду. Архимед вычислил «квадратуру параболы», т.е. подсчитал площадь фигуры, ограниченной частью квадратичной параболы (рис. 301). Архимед разбил площадь параболического сегмента на треугольники. При этом он подсчитал, что площадь каждого нового треугольника в 4 раза меньше площади предыдущего. Зная площадь первого треугольника $a_1 = 1$, они пришел к необходимости подсчитать бесконечную сумму

$$1+\frac{1}{4}+\frac{1}{4^2}+\frac{1}{4^3}+\ldots$$

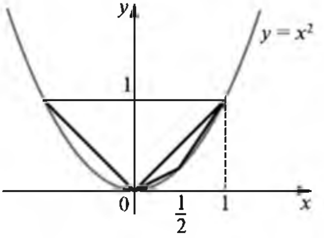


Рис. 301

Нарисовав квадратики с уменьшающимися вдвое сторонами (и, следовательно, вчетверо уменьшающимися площадями), он увидел, что, складывая по три четверти этих квадратиков до бесконечности, он получит площадь всего квадрата. Чтобы $\frac{3}{4}$ площади первого квадрата равнялись бы площади первого треугольника, вписанного в параболический сегмент, надо, чтобы $\frac{3}{4}a^2 = 1$, т.е. чтобы площадь исходного квадрата a^2 равнялась бы $\frac{4}{3}$. Это и будет площадь параболического сегмента. Архимед произвел «квадратуру» — построил квадрат, равновеликий данной фигуре.

Фактически он нашел сумму бесконечной геометрической прогрессии с первым членом 1 и знаменателем $\frac{1}{4}$: $1+\frac{1}{4}+\frac{1}{4^2}+\ldots=\frac{4}{3}$.

Пример

Найти сумму s ряда $\frac{1}{1\cdot 2}+\frac{1}{2\cdot 3}+\frac{1}{3\cdot 4}+\ldots$.

Частичную сумму этого ряда мы вычислили ранее:

$$s_n = \frac{1}{1\cdot 2}+\frac{1}{2\cdot 3}+\ldots+\frac{1}{n(n+1)}=1-\frac{1}{n+1}.$$

Ясно, что с ростом n дробь $\frac{1}{n+1}$ неограниченно приближается к нулю, а сумма s_n приближается к числу 1, которое и является суммой ряда.

Ответ: $s = 1$.

Геометрическую прогрессию называют бесконечно убывающей, если ее знаменатель q по модулю меньше единицы: $|q| < 1$.

Такое название возникло потому, что при $|q| < 1$ общий член прогрессии $a_n = a_1 q^{n-1}$ становится сколь угодно малым, «бесконечно убывает».

Найдем сумму бесконечно убывающей геометрической прогрессии с первым членом a_1 и знаменателем q ($|q| < 1$).

Вычислим сумму n ее членов:

$$s_n = a_1 + a_1 q + \ldots + a_1 q^{n-1} = a_1 \cdot \frac{1-q^n}{1-q}.$$

Разобьем s_n на два слагаемых:

$$s_n = \frac{a_1}{1-q} - a_1\frac{q^n}{1-q}.$$

Первое слагаемое постоянно, а второе «бесконечно уменьшается» с ростом n.

При сложении «до бесконечности» мы отбрасываем эту добавку и получаем формулу

$$s = \frac{a_1}{1-q}.$$

Проверяя эту формулу при $a_1 = 1$, $q = \frac{1}{4}$, получаем результат Архимеда: площадь параболического сегмента с основанием 2 и высотой 1 равно $\frac{4}{3}$.

С помощью формулы суммы бесконечно убывающей геометрической прогрессии легко обращать периодические десятичные дроби в обыкновенные.

Примеры

1. Обратить дробь $a = 1{,}444... = 1{,}(4)$ в обыкновенную. Число $1{,}444...$ — это запись суммы $1 + 0{,}4 + 0{,}04 + 0{,}004 + ...$ Сумму $0{,}4 + 0{,}04 + ...$ можно рассмотреть как сумму бесконечно убывающей геометрической прогрессии с первым членом $a_1 = 0{,}4$ и знаменателем $q = 0{,}1$. Выполняем подсчеты:

$$a = 1 + \frac{a_1}{1-q} = 1 + \frac{0{,}4}{1-0{,}1} = 1 + \frac{4}{9} = \frac{13}{9}.$$

2. $a = 2{,}3(12) = 2{,}312121212...$ Запишем равенство:

$$a = 2{,}3 + 0{,}012 + 0{,}00012 + ... = 2{,}3 + 0{,}012\,(1 + 0{,}01 + 0{,}01^2 + ...).$$

В скобках стоит бесконечно убывающая геометрическая прогрессия со знаменателем $q = 0{,}01$. Применяем формулу:

$$a = 2 + \frac{3}{10} + \frac{12}{1000} \cdot \frac{1}{0{,}99} = 2 + \frac{3}{10} + \frac{4}{10 \cdot 33} = 2 + \frac{103}{330} = \frac{763}{330}.$$

Теория пределов помогла обосновать многие рассуждения о суммировании рядов, однако она не устранила всех противоречий, которые возникают, если «легкомысленно» пользоваться этим понятием.

Рассмотрим, например, такой ряд:

$$1 - 1 + 1 - 1 + ...$$

Его частичные суммы попеременно равны то единице (при нечетных n), то нулю (при четных n):

$$1 = 1,\ 1 - 1 = 0,\ 1 - 1 + 1 = 1 \text{ и т.д.}$$

Поэтому наш ряд не имеет суммы в смысле обычной теории пределов. Но может быть, можно его сумму «сосчитать» другим способом.

Итак, пусть $S = 1 - 1 + 1 - 1 + \ldots$

Сделаем преобразования:

$$S = (1 - 1) + (1 - 1) + \ldots = 0 + 0 + \ldots = 0.$$

Поступим иначе:

$$S = 1 - (1 - 1) - (1 - 1) - \ldots = 1 - 0 - 0 - \ldots = 1.$$

А если так:

$$S = 1 - (1 - 1 + 1 - \ldots) = 1 - S \Rightarrow 2S = 1 \Rightarrow S = \frac{1}{2}.$$

Так чему же равно S — нулю, единице или половине? Ответ на этот вопрос зависит от определений, соглашений, принимаемых нами о том, что мы понимаем под суммой бесконечного ряда $1 - 1 + \ldots$ При некоторых разумных и весьма полезных соглашениях получается как раз последний ответ $S = \frac{1}{2}$, кажущийся самым нелепым.

Контрольные вопросы и задания

1. Что такое числовая последовательность?
2. Какие вы знаете способы задания последовательности?
3. Какими рекуррентными соотношениями задаются прогрессии?
4. Какие свойства последовательностей вы можете определить?
5. Как задается последовательность чисел Фибоначчи?
6. В чем состоит метод математической индукции?
7. Что такое предел последовательности?
8. Какой признак сходимости последовательности вы знаете?
9. Приведите геометрические примеры сходящихся последовательностей.
10. Что такое ряд и его сумма?
11. Что такое бесконечно убывающая геометрическая прогрессия?
12. Как найти сумму бесконечно убывающей геометрической прогрессии?
13. Как записать периодическую десятичную дробь в виде суммы постоянного числа и бесконечно убывающей геометрической прогрессии?

§ 30. Производная и ее вычисление

Механический смысл производной

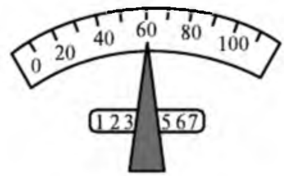


Рис. 302

Производная — это скорость. Представим себе, что мы едем в кабине автомашины по прямолинейному шоссе. Посмотрим на счетчик километража. Запомнив показания счетчика в начальный момент, можно определить проделанный путь в любой момент времени. Посмотрим на стрелку спидометра. Она покажет скорость движения в данный момент (рис. 302). Так с нашим движением связаны две величины — путь s и скорость v, которые являются функциями времени t: $s = s(t)$, $v = v(t)$.

Функции s и v связаны между собой. В конце XVII в. великий английский ученый Исаак Ньютон открыл общий способ описания этой связи. С помощью этого способа для каждой функции s можно построить новую функцию v. Эту функцию называют *производной функции* s, а сам переход от функции s к функции v — дифференцированием.

Открытие Ньютона явилось поворотным пунктом в истории естествознания. Оказалось, что количественные характеристики самых различных процессов, исследуемых в физике, химии, биологии, в технических дисциплинах, могут быть выражены на языке математического анализа, изучающего связи между функциями и их производными.

Самой важной и самой простой моделью для построения производной остается изученная еще Ньютоном модель механического движения. В этой модели исходной функцией является путь, ее производной — скорость. Вот почему

самым простым и коротким ответом на вопрос, что такое производная, является такой: производная — это скорость.

А что такое скорость? Оказывается, на этот простой вопрос ответить не так просто. Прочтите диалог между водителем (женщиной) и полицейским, остановившим автомобиль за превышение скорости. Он взят из знаменитых «Фейнмановских лекций по физике»:

— Мадам, Вы нарушили правила уличного движения. Вы ехали со скоростью 90 км/ч.

— Простите, это невозможно. Как я могла проехать 90 км/ч, если я еду всего лишь 7 минут!
— Я имею в виду, мадам, что если бы вы продолжили ехать таким же образом, то через час вы бы проехали 90 километров.
— Если бы я продолжала ехать, как ехала, еще час, то налетела бы на стенку в конце улицы!
— Ваш спидометр показывал 90 км/ч.
— Мой спидометр сломан и давно не работает.
Вы видите, что полицейский не смог объяснить, что такое скорость 90 км/ч. А вы смогли бы?
Попробуйте объяснить, что такое скорость равномерного движения и как ее можно измерить.

Разберемся в том, что же такое скорость произвольного движения.

Средняя и мгновенная скорости

Пусть точка движется по прямой и задан закон, по которому можно вычислить путь s как функцию времени t:

$$s = s(t).$$

Например, если точка движется равномерно со скоростью v, то $s = vt$ (считаем, что при $t = 0$ путь $s = 0$). Если точка движется под действием силы тяжести с нулевой начальной скоростью, то $s = \frac{gt^2}{2}$ (мы считаем, что g — ускорение свободного падения — постоянно). Возможны и другие законы движения. Так, тело, запущенное с поверхности Земли вертикально вверх с начальной скоростью, позволяющей полностью преодолевать земное притяжение (вторая космическая скорость), удаляется от центра Земли по закону $s = A(t + C)^{2/3}$, где A и C — некоторые константы.

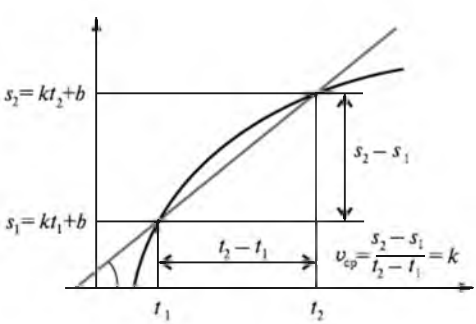


Рис. 303

Определим среднюю скорость точки за промежуток времени $[t_1, t_2]$, как отношение пройденного пути ко времени движения (рис. 303):

$$\upsilon_{cp} = \frac{s(t_2) - s(t_1)}{t_2 - t_1}.$$

Определим скорость точки в момент времени t_1

(ее в механике называют мгновенной скоростью; рис. 304). Для этого вычислим среднюю скорость за промежуток времени $[t_1, t_2]$. Уменьшая отрезок $[t_1, t_2]$, т.е. приближая t_2 к t_1, мы заметим, что значение средней скорости при этом практически не меняется, а вернее, приближается к некоторому числу, которое и считается значением скорости в момент времени t_1.

Рис. 304

Пример

Длина пути задается функцией $s = \frac{gt^2}{2}$. Зафиксируем произвольный момент времени и вычислим среднюю скорость на отрезке $[t; t_1]$:

$$\upsilon_{ср} = \frac{s(t_1)-s(t)}{t_1-t} = \frac{gt_1^2/2-gt^2/2}{t_1-t} = \frac{g}{2}\frac{t_1^2-t^2}{t_1-t} = \frac{g}{2}(t_1+t).$$

Если стягивать отрезок $[t; t_1]$ к точке t, т.е. брать значения t_1 все ближе и ближе к t, то сумма $t_1 + t$ будет приближаться к $t + t = 2t$, а все выражение $\frac{g}{2}(t_1 + t)$ — к $\frac{g}{2}2t = gt$. Это и есть мгновенная скорость в точке t: $v = gt$. Переход от средней скорости на отрезке $[t; t_1]$ к мгновенной скорости в точке t при стягивании отрезка $[t; t_1]$ к точке t называется ***предельным переходом***. Обычно говорят так: при стремлении t_1 к t выражение $\frac{g}{2}(t_1 + t)$ стремиться к gt. Запишем это:

$$\lim_{t_1 \to t}\frac{g}{2}(t_1+t) = gt.$$

Используя слово «предел», можно сказать, что мгновенная скорость в точке t — это предел средней скорости при стягивании отрезка, на котором измеряется эта скорость, к точке t.

Средняя скорость вычисляется для каждого отрезка времени, мгновенная скорость вычисляется в каждый момент времени.

Средняя скорость является приближением к мгновенной, и это приближение тем точнее, чем меньше отрезок.

Геометрический смысл производной

Производная — это угловой коэффициент касательной. Одновременно с Ньютоном к задаче нахождения производной пришел немецкий математик Г. Лейбниц, изучая касательные к произвольным кривым.

Наглядное представление о том, как провести касательную, есть у каждого человека. Надо вообразить себе, что кривая изготовлена из жесткого материала (например, из проволоки), и вы приставляете линейку так, чтобы она коснулась кривой в выбранной точке (рис. 305).

Если вы вырезаете из бумаги криволинейную фигуру, то ножницы направлены по касательной к ее границе.

Постараемся перевести наглядное представление о касательной на более точный язык. Будем считать, что кривая — это ломаная с очень большим числом маленьких звеньев. Именно такая точка зрения была у создателей дифференциального исчисления. В первом учебнике по анализу, написанном 300 лет назад последователем Лейбница маркизом Лопиталем, дано следующее определение касательной: «Если продолжить одно из маленьких звеньев ломаной, составляющих кривую линию, то эта продолженная таким образом сторона будет называться *касательной к кривой*» (рис. 306).

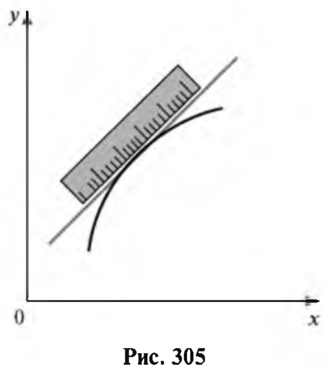


Рис. 305

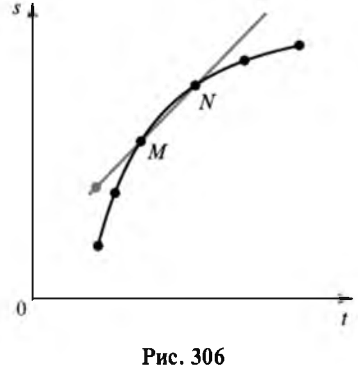


Рис. 306

Хорошее представление о касательной дает следующее описание. Посмотрим в микроскоп на маленький участок кривой. На рисунке 307 в разных масштабах изображены участки одной и той же параболы около точки M. На первом графике (рис. 307, *а*) видно, что парабола искривлена, на втором (рис. 307, *б*) это искривление уже менее заметно, а на третьем (рис. 307, *в*) — участок кривой почти неотличим от отрезка прямой линии, которая и является касательной к параболе в точке M.

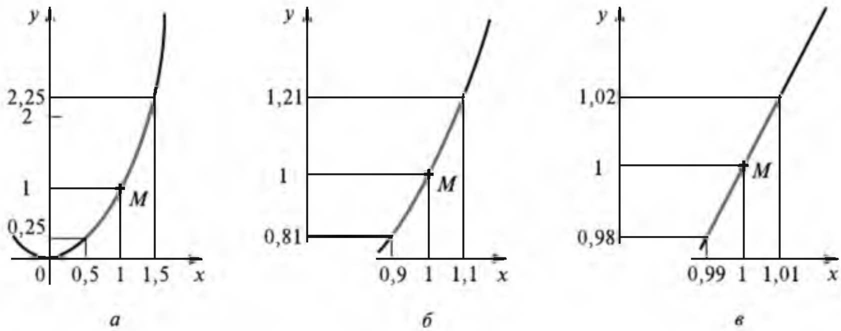


Рис. 307

Более точно объяснить, что такое касательная, так же не просто, как и дать точное определение скорости. Для этого понадобится предельный переход, аналогичный тому, который мы совершили при определении скорости.

Пусть дана некоторая кривая и точка *P* на ней (рис. 308). Возьмем на этой кривой другую точку P_1 и проведем прямую через точки *P* и P_1. Эту прямую обычно называют секущей. Станем приближать точку P_1 к *P*. Положение секущей PP_1 будет меняться, но с приближением P_1 к *P* начнет стабилизироваться. Предельное положение секущей PP_1 при стремлении точки P_1 к точке *P* и будет касательной к кривой в точке *P*.

Как перевести описанное построение на язык формул? Пусть кривая является графиком функции $y = f(x)$, а точка *P* задана своими координатами (x, y). Касательная является некоторой прямой, проходящей через точку *P*. Чтобы построить эту прямую, достаточно найти ее угловой коэффициент.

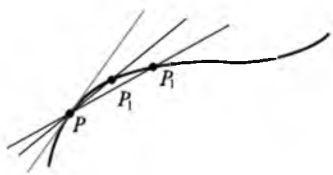


Рис. 308

Обозначим угловой коэффициент касательной *k*. Сначала найдем угловой коэффициент k_1 секущей PP_1. Пусть абсцисса точки P_1 равна x_1. Из рисунка 309 ясно, что $k_1 \frac{f(x_1)-f(x)}{x_1-x}$. Для нахождения *k* надо устремить x_1 к x. Тогда точка P_1 начнет приближаться к точке *P*, а секущая PP_1 — к касательной

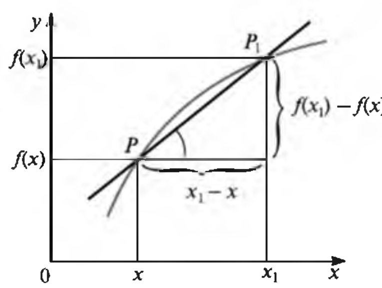


Рис. 309

в точке P. Таким образом, угловой коэффициент касательной можно найти как предел выражения $\frac{f(x_1)-f(x)}{x_1-x}$ при стремлении x_1 к x, или, в символической записи, $k(x)=\lim\limits_{x_1 \to x}\frac{f(x_1)-f(x)}{x_1-x}$.

Мы пришли к той же задаче, которая встретилась при нахождении скорости — осуществить предельный переход в выражении вида $\frac{f(x_1)-f(x)}{x_1-x}$ при стремлении x_1 к x. Этот предельный переход называется *дифференцированием функции* f.

Дифференцирование, или нахождение производной, — это новая математическая операция, имеющая тот же смысл, что в механике — нахождение скорости, а в геометрии — вычисление углового коэффициента касательной.

Для нахождения производной в данной точке надо рассмотреть маленький участок изменения аргумента около этой точки. Производная приближенно равна средней скорости на этом участке (на языке механики) или угловому коэффициенту секущей (на языке геометрии). Для точного вычисления производной надо совершить предельный переход — стянуть отрезок изменения аргумента в точку. Тогда средняя скорость превратится в мгновенную, а секущая — в касательную, и мы вычислим производную.

Определение производной

Математический анализ, созданный Ньютоном и Лейбницем, долго развивался на основе интуитивного понятия производной как «скорости изменения функции». Современное определение производной появилось лишь в XIX в. после того, как были уточнены основные понятия математического анализа: вещественное число, функция, предел. С их помощью можно дать следующее определение производной.

Производной функции $y = f(x)$ называется предел отношения $\frac{f(x_1)-f(x)}{x_1-x}$ при стремлении x_1 к x.

Исторически сложилась символика для обозначения участвующих в этом определении выражений. Разность значений аргумента обозначают двойной буквой Δx («дельта икс») и называют приращением аргумента, а разность значений функций — Δy — приращением функции.

Иначе говоря, разность $x_1 - x$ обозначается Δx (и тогда $x_1 = x + \Delta x$), а разность $f(x + \Delta x) - f(x)$ обозначается Δy. Средняя скорость изменения функции записывается так: $\frac{\Delta y}{\Delta x}$. Стягивание отрезка в точку означает стремление Δx к нулю (рис. 310). Производную функции $y = f(x)$ обозначают с помощью штриха: y' или f'. Получается новый вариант определения производной.

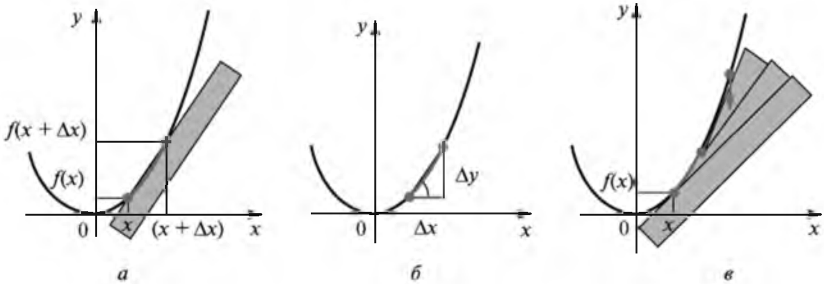


Рис. 310

Производной называется предел отношения приращения функции к приращению аргумента, когда приращение аргумента стремится к нулю.

Символически определение производной можно записать так:

$$y' = \lim_{\Delta x \to 0} \frac{\Delta y}{\Delta x} \text{ или } \frac{\Delta y}{\Delta x} \underset{\Delta x \to 0}{\to} y'.$$

Как поступают при вычислении производной?

Первый шаг: вычисляют Δy — приращение функции на отрезке $[x, x + \Delta x]$ — и составляют отношение $\frac{\Delta y}{\Delta x}$.

Второй шаг: полагают, что Δx приближается к нулю, и находят, к какому значению приближается отношение $\frac{\Delta y}{\Delta x}$ при $\Delta x \to 0$.

Предельные переходы

При определении скорости, при нахождении углового коэффициента касательной, при вычислении производной мы совершаем предельные переходы. Попробуем более точно описать их суть.

Мы рассматривали переменные величины. Пусть, например, переменная y является функцией переменной величины x. Нас интересует поведение y при приближении аргумента x к некоторому значению a. Отметим, что значение a может входить в область определения функции, а может и не входить. Так, исследуя функции $y_1 = 3 + x^2$, $y_2 = \dfrac{1-x^2}{1-x}$, $y_3 = \dfrac{\sqrt{x}-1}{x^2-1}$, $y_4 = \dfrac{x}{x+1}$ вблизи значения $x = 1$, замечаем, что это значение можно подставить в формулы для y_1 и y_4 и нельзя в формулы для y_2 и y_3. Предельный переход для функций y_1 и y_4 при приближении x к 1 (при $x \to 1$) совершить просто — надо вычислить значения этих функций при $x = 1$:

$$\lim_{x\to 1} y_1 = \lim_{x\to 1} (3 + x^2) = 3 + 1 = 4; \ \lim_{x\to 1} y_4 = \lim_{x\to 1} \frac{x}{x+1} = \frac{1}{2}.$$

Так обычно ведут себя все функции, заданные известными нам формулами:

предел функции при стремлении аргумента к значению a, входящему в область определения, равен значению функции в точке a: $\lim\limits_{x\to 1} f(x) = f(a)$, если a входит в область определения функции f.

Этот принцип, который отражает непрерывность всех известных нам функций вблизи тех точек, где они определены, позволяет осуществить большое количество важных предельных переходов. Мы так и будем называть его: *принцип непрерывности*.

Принцип непрерывности удобно формулируется на языке приращений: если $\Delta x \to 0$, то $\Delta y \to 0$.

Предельные переходы для функций y_2 и y_3 при $x \to 1$ так просто осуществить нельзя: y_2 и y_3 не определены при $x = 1$. То, что мы хотим сделать с этими функциями, поясним на примере функции $y_5 = \dfrac{x}{x}$. Ясно, что при всех $x \neq 0$ значения этой функции равны 1. При $x = 0$ эта функция не определена, но очевидно, что ее предельным значением при $x \to 0$ надо считать $y = 1$. Это значение получается при сокращении числителя и знаменателя дроби $\dfrac{x}{x}$ на x. Аналогичное сокращение можно произвести и для функций y_2 и y_3:

$$y_2 = \frac{1-x^2}{1-x} = \frac{(1-x)(1+x)}{1-x} = 1+x;$$

$$y_3 = \frac{\sqrt{x}-1}{x^2-1} = \frac{\sqrt{x}-1}{(x-1)(x+1)} = \frac{\sqrt{x}-1}{(\sqrt{x}-1)(\sqrt{x}+1)(x+1)} = \frac{1}{(\sqrt{x}+1)(x+1)}.$$

Полученные после сокращения выражения уже определены при $x = 1$ и поэтому предельные значения можно получить, подставляя в них число 1 вместо x:

$$\lim_{x\to 1} y_2 = \lim_{x\to 1} (1 + x) = 1 + 1 = 2;$$

$$\lim_{x\to 1} y_3 = \lim_{x\to 1} \frac{1}{(\sqrt{x}+1)(x+1)} = \frac{1}{(1+1)(1+1)} = \frac{1}{4}.$$

Приведенные примеры не исчерпывают всего разнообразия встречающихся случаев — в дальнейшем мы будем рассматривать и более сложные предельные переходы.

Контрольные вопросы и задания

1. Обратите внимание на следующие ключевые слова, встретившиеся в этом параграфе: средняя скорость, мгновенная скорость, секущая, касательная, приращение аргумента, приращение функции, производная. Приведите примеры их использования.
2. Какое движение называется равномерным?
3. Как вычислить скорость равномерного движения?
4. Что такое средняя скорость неравномерного движения?
5. Как вычислить мгновенную скорость точки?
6. Чем отличается мгновенная скорость от средней скорости?
7. Что можно сказать о движении, если его средняя скорость одна и та же на любом отрезке времени?
8. Что такое касательная?
9. Как вычислить угловой коэффициент секущей, проходящей через две точки графика некоторой функции?
10. Как найти угловой коэффициент касательной?
11. Для какой функции касательная к ее графику совпадает с самим графиком?
12. Что такое производная?
13. Как вычислить производную?
14. Что такое производная с механической точки зрения?
15. Что такое производная с геометрической точки зрения?

§ 31. Вычисление производной

Схема вычисления производной

Производные вычисляем по следующей схеме:

1. Находим приращение функции $y = f(x)$ на промежутке $[x; x + \Delta x]$:
$$\Delta y = f(x + \Delta x) - f(x).$$
2. Делим приращение функции на приращение аргумента, т.е. находим среднюю скорость роста функции f на отрезке $[x; x + \Delta x]$:
$$v_{\text{ср}} = \frac{\Delta y}{\Delta x} = \frac{f(x+\Delta x)-f(x)}{\Delta x}.$$
3. Переходим к пределу, т.е. стараемся преобразовать среднюю скорость $\frac{\Delta y}{\Delta x}$ так, чтобы можно было легко увидеть мгновенную скорость — функцию, значение которой в точке x близко к $\frac{\Delta y}{\Delta x}$ при малых Δx. В этих преобразованиях часто удается сократить Δx.

Примеры

1. Производная постоянной функции $y = C$.
Функция определена на всей вещественной оси $\mathbb{R}$ и принимает постоянное значение C в каждой точке, поэтому ее приращение
$\Delta y = C - C = 0 \Rightarrow v_{\text{ср}} = \frac{\Delta y}{\Delta x} = 0.$
Так как отношение $\frac{\Delta y}{\Delta x}$ постоянно и равно нулю, то производная y тоже равна нулю: $y' = 0$. Итак,

производная постоянной равна нулю (рис. 311).

Рис. 311

2. Производная линейной функции $y = ax + b$.
   Функция определена на всей вещественной оси R, и ее приращение равно

$$\Delta y = a(x + \Delta x) + b - (ax + b) = ax + a\Delta x + b - ax - b =$$
$$= a\,\Delta x \Rightarrow \frac{\Delta y}{\Delta x} = \frac{a\,\Delta x}{\Delta x} = a.$$

   Отношение $\frac{\Delta y}{\Delta x}$ является постоянным. Поэтому в качестве производной надо взять функцию, принимающую это постоянное значение a: $y' = a$. Итак,

производная линейной функции равна коэффициенту при переменной.

   Полученный результат неудивителен — линейная функция соответствует равномерному движению, скорость которого постоянна (рис. 312).
3. Производная функции $y = x^3$.
   Вычислим Δy:

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

   Найдем отношение: $\frac{\Delta y}{\Delta x} = 3x^2 + 3x\Delta x + (\Delta x)^2$.
   При малых значениях Δx величина дроби $\frac{\Delta y}{\Delta x}$ определяется только первым слагаемым. В качестве производной надо взять функцию $g(x) = 3x^2$. Получаем $y' = 3x^2$ (рис. 313).

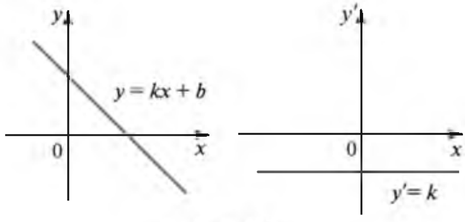


Рис. 312

4. Производная функции $y = \frac{1}{x}$.
   Эта функция определена при всех x, кроме $x = 0$. Найдем отношение приращения функции к приращению аргумента:

$$\Delta y = \frac{1}{x + \Delta x} - \frac{1}{x} = -\frac{\Delta x}{x(x + \Delta x)}\ ;\ \frac{\Delta y}{\Delta x} = -\frac{1}{x(x + \Delta x)}.$$

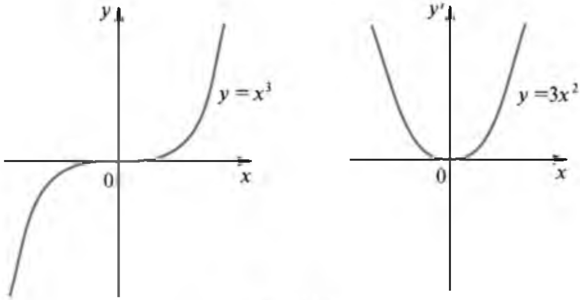


Рис. 313

При малых значениях Δx выражение $x + \Delta x$ близко к x, а вся дробь $-\dfrac{1}{x(x+\Delta x)}$ мало отличается от дроби $-\dfrac{1}{xx}=-\dfrac{1}{x^2}$. Следовательно, $y' = -\dfrac{1}{x^2}$ (рис. 314).

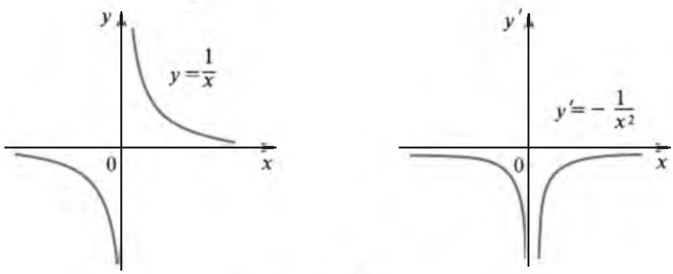


Рис. 314

Эти примеры затрагивали только многочлены и дроби — рациональные функции. При вычислении производной рациональной функции дробь $\dfrac{\Delta y}{\Delta x}$ всегда сократится на Δx. После этого можно подставить $\Delta x = 0$ и получить производную.

5. Производная функции $y=\sqrt{x}$.
Эта функция определена при всех $x \geq 0$:

$$\Delta y=\sqrt{x+\Delta x}-\sqrt{x}\,,\quad \frac{\Delta y}{\Delta x}=\frac{\sqrt{x+\Delta x}-\sqrt{x}}{\Delta x}.$$

Для того чтобы сократить дробь на Δx, умножим ее числитель и знаменатель на сумму радикалов:

$$\frac{\Delta y}{\Delta x}=\frac{\sqrt{x+\Delta x}-\sqrt{x}}{\Delta x}\cdot\frac{\sqrt{x+\Delta x}+\sqrt{x}}{\sqrt{x+\Delta x}+\sqrt{x}}=\frac{x+\Delta x-x}{\Delta x(\sqrt{x+\Delta x}+\sqrt{x})}=$$

$$=\frac{\Delta x}{\Delta x(\sqrt{x+\Delta x}+\sqrt{x})}=\frac{1}{\sqrt{x+\Delta x}+\sqrt{x}}.$$

При малых Δx значение корня $\sqrt{x+\Delta x}$ приближается к $\sqrt{x}$. Поэтому при переходе к пределу надо заменить $\sqrt{x+\Delta x}$ в знаменателе на $\sqrt{x}$:

$$y'=\frac{1}{\sqrt{x}+\sqrt{x}}=\frac{1}{2\sqrt{x}}\text{ (рис. 315).}$$

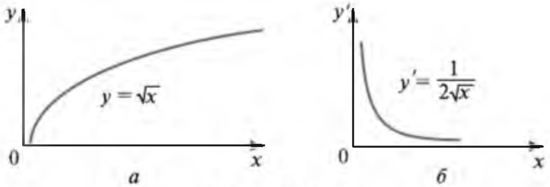


Рис. 315

Правила дифференцирования

1. Производная суммы

Производная суммы двух функций равна сумме их производных:

$$(u+v)'=u'+v'.$$

2. Вынесение постоянной за знак производной

Если функцию умножить на постоянное число, то и производная умножится на это число:

$$(Cy)'=Cy'.$$

3. Производная произведения

$$(uv)'=u'v+uv'.$$

4. Производная частного

$$\left(\frac{u}{v}\right)'=\frac{u'v-uv'}{v^2}.$$

Доказательство. 1. Пусть функция w равна сумме функций u и v, т.е. $w(x)=u(x)+v(x)$. Вычислим $\Delta w=w(x+\Delta x)-w(x)=u(x+\Delta x)+v(x+\Delta x)-[u(x)+v(x)]=u(x+\Delta x)-u(x)+v(x+\Delta x)-v(x)=\Delta u+\Delta v$.

Составим дробь $\frac{\Delta w}{\Delta x}$:

$$\frac{\Delta w}{\Delta x}=\frac{\Delta u+\Delta \upsilon}{\Delta x}=\frac{\Delta u}{\Delta x}+\frac{\Delta \upsilon}{\Delta x}.$$

Средняя скорость роста суммы двух функций равна сумме средних скоростей слагаемых. При переходе к пределу это соотношение должно сохраниться: если $\frac{\Delta u}{\Delta x}$ близко к u', а $\frac{\Delta \upsilon}{\Delta x}$ близко к v', то $\frac{\Delta u}{\Delta x}+\frac{\Delta \upsilon}{\Delta x}$ близко к $u' + v'$. Итак, $\frac{\Delta w}{\Delta x} \to u' + v'$, т.е.

$$(u + v)' = u' + v'.$$

2. Вычислим приращение функции Cy:

$$\Delta(Cy) = Cy(x + \Delta x) - Cy(x) = C\,\Delta y, \quad \frac{\Delta(Cy)}{\Delta x}=C\frac{\Delta y}{\Delta x}.$$

При умножении функции на постоянное число средняя скорость также умножится на это число. Это соотношение сохранится и в пределе, т.е. если $\frac{\Delta y}{\Delta x}$ близко к y', то $C\frac{\Delta y}{\Delta x}$ близко к Cy'.

Итак,

$$\frac{(Cy)}{x} \to Cy', \text{ т.е. } (Cy)' = Cy'.$$

3. Вычислим среднюю скорость изменения произведения.
Пусть $y = uv$. Тогда $\Delta y = u(x + \Delta x)\, v\,(x + \Delta x) - u(x)\, v(x)$. Заменим $u(x + \Delta x)$ на $u(x) + \Delta u$ и $v(x + \Delta x)$ на $v(x) + \Delta v$. Получим:

$$\Delta y = (u + \Delta u)\,(v + \Delta v) - uv = u\,\Delta v + \Delta uv + \Delta u\,\Delta v;$$
$$\Delta y = \Delta uv + u\,\Delta v + \Delta u\,\Delta v;$$
$$\frac{\Delta y}{\Delta x}=\frac{\Delta u}{\Delta x}\upsilon+u\frac{\Delta \upsilon}{\Delta x}+\frac{\Delta u}{\Delta x}\Delta\upsilon.$$

Пусть $\Delta x \to 0$. Тогда $\frac{\Delta u}{\Delta x} \to u'$, $\frac{\Delta \upsilon}{\Delta x} \to v'$. Сделаем предельный переход в каждом слагаемом. Если $\frac{\Delta u}{\Delta x} \to u'$, то $\frac{\Delta u}{\Delta x}v \to u'v$. Если $\frac{\Delta \upsilon}{\Delta x} \to v'$, то $u\frac{\Delta \upsilon}{\Delta x} \to uv'$. Третье слагаемое является произведением двух переменных множителей, ведущих себя по-разному при $\Delta x \to 0$. Мы уже знаем, что $\frac{\Delta u}{\Delta x} \to u'$. Если $\Delta x \to 0$, то, по принципу непрерывности, $\Delta v \to 0$.

Собирая все слагаемые вместе, получаем

$$\frac{\Delta y}{\Delta x} \to u'v + uv',$$

что и требовалось доказать.

* Правило 2 можно получить как следствие правила 3, считая, что Cy есть произведение постоянной функции $u = C$ на $v = y$:

$$(Cy)' = (C')y + C(y)' = 0 \cdot y + Cy' = Cy'.$$

Переход к пределу, который совершается при доказательстве правила 2, гораздо проще того, который нужен для правила 3, поэтому мы дали отдельное доказательство важного правила дифференцирования — вынесение постоянного числа за знак производной.

4. Вычисления можно провести по стандартной схеме, находя среднюю скорость частного. Мы поступим иначе. Обозначим $\frac{u}{\upsilon}$ через h. Получим $\frac{u}{\upsilon} = h$, $u = \upsilon h$. Найдем производную функции u по правилу дифференцирования произведения:

$$u' = \upsilon' h + \upsilon h'.$$

Выразим из этой формулы h', а вместо h подставим его значение $\frac{u}{\upsilon}$:

$$h' = \frac{u' - \upsilon' h}{\upsilon} = \frac{u' - \upsilon'\left(\frac{u}{\upsilon}\right)}{\upsilon} = \frac{u'\upsilon - \upsilon' u}{\upsilon^2}.$$

Окончательно

$$\left(\frac{u}{\upsilon}\right)' = \frac{u'\upsilon - u\upsilon'}{\upsilon^2}.$$

Производную функции $y = \frac{1}{x}$ можно получить, пользуясь правилом 4. Представим $y = \frac{1}{x}$ как частное функций $u = 1$ и $v = x$. Тогда

$$\left(\frac{1}{x}\right)' = \frac{(1)'(x) - 1 \cdot (x)'}{x^2} = \frac{0 - 1}{x^2} = -\frac{1}{x^2}.$$

Производная степенной функции

Производную любой степени с натуральным показателем можно получить, пользуясь правилом дифференцирования произведения.

Пример

Найдем производную функции $y = x^4$.

Представим x^4 как x^3x. Зная производные от $y = x^3$ и $y = x$, вычислим производную от произведения:

$$(x^4)' = (x^3)'x + (x^3)\,x' = 3x^2x + x^3 \cdot 1 = 4x^3.$$

Сопоставим друг с другом формулы $x' = 1$; $(x^2)' = 2x$; $(x^3)' = 3x^2$; $(x^4)' = 4x^2$.

Общая закономерность: $(x^n)' = nx^{n-1}$.

Проверить это можно так. Предположим, что $(x^{n-1})' = (n-1)\,x^{n-2}$. Представив $x^n = x^{n-1}x$, найдем

$$(x^n)' = (x^{n-1})'x + x^{n-1}(x)' = (n-1)x^{n-2}x + x^{n-1} =$$
$$= (n-1)x^{n-1} + x^{n-1} = nx^{n-1}.$$

Таким образом, если формула $(x^k)' = k \cdot x^{k-1}$ верна при $k = n - 1$, то она верна и при $k = n$. Нам эта формула уже известна, например, при $k = 2$. Тогда она верна и при $k = 3$, а следовательно, и при $k = 4$, $k = 5$ и т.д., т.е. при любом натуральном k.

Напомним, что рассуждение, которое мы здесь провели, называется *математической индукцией*.

Теперь найдем производную функции $y = \dfrac{1}{x^n}$:

$$y' = \frac{0 - nx^{n-1}}{(x^n)^2} = -\frac{n}{x^{2n-(n-1)}} = -\frac{n}{x^{n+1}}.$$

Если дробь $y = \dfrac{1}{x^n}$ записать как $y = x^k$, где $k = -n$ (т.е. как степень с отрицательным показателем), то $y' = (-n) \cdot x^{-n-1}$. Заменив $(-n)$ на k, получим $y' = kx^{k-1}$, т.е. формула дифференцирования степени, выведенная нами для натуральных показателей, остается верной и для целых отрицательных показателей: $(x^k)' = kx^{k-1}$, где k — любое целое число, кроме нуля. На самом деле выведенная формула остается верной и для дробных показателей, которые мы будем изучать далее.

Примеры

1. $y = 2x^3 + 3x + 1.$

$$y' = (2x^3)' + (3x)' + 1' = 6x^2 + 3.$$

2. $y = \dfrac{2x^2 - 1}{x}.$

Преобразуем исходную функцию к сумме, выполним почленное деление: $y = \dfrac{2x^2-1}{x} = 2x + \dfrac{-1}{x}$. Теперь производная вычисляется аналогично предыдущему случаю:

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

3. $y = \sqrt{3x} + 1.$

$$y' = (\sqrt{3}\sqrt{x})' + 1' = \frac{\sqrt{3}}{2\sqrt{x}} + 0 = \frac{\sqrt{3}}{2\sqrt{x}}.$$

4. $y = \dfrac{x}{x^2+1}.$

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

5. $y = \dfrac{x^{200}-1}{x^{100}}.$

$$y = \frac{x^{200}-1}{x^{100}} = x^{100} - \frac{1}{x^{100}} = x^{100} - x^{-100}.$$

$$y' = 100x^{99} + 100x^{-101} = 100\,\frac{x^{200}+1}{x^{101}}.$$

Изменение производной при линейной замене аргумента

Запишем следующие полезные формулы дифференцирования:

1. $[f(x-a)]' = f'(x-a);$
2. $[f(kx)]' = kf'(kx).$

Эти формулы можно объединить в одну:

$$[f(kx+b)]' = kf'(kx+b).$$

Физический смысл приведенных формул очень простой. Переход от функции $y = f(x)$ к функции $y = f(kx + b)$ означает изменение начала отсчета и масштаба переменной x. Изменение начала отсчета не может повлиять на вычисление скорости — показания спидометра не зависят от начального показания счетчика километража. Изменение масштаба измерения аргумента (времени) в 6 раз повлечет за собой такое же изменение скорости. Так, скорость, измеренная в километрах в час, в 60 раз меньше скорости, измеренной в километрах в минуту.

Доказательство. Вычислим производную функции $y = f(kx + b)$, следуя общей схеме вычисления производной. Предварительно введем обозначения. Функцию $y = f(kx + b)$ как функцию от x обозначим через g: $g(x) = f(kx + b)$. Выражение $kx + b$ обозначим через z: $z = kx + b$, т.е. $g(x) = f(z)$. Свяжем между собой Δx и Δz: $\Delta z = k(x + \Delta x) + b - (kx + b) = k\,\Delta x$, т.е. $\Delta z = k\,\Delta x$. Вычислим $\frac{\Delta g}{\Delta x}$:

$$\frac{\Delta g}{\Delta x} = \frac{g(x+\Delta x)-g(x)}{\Delta x} = \frac{f(z+\Delta z)-f(z)}{\Delta x} = \frac{f(z+\Delta z)-f(z)}{\Delta z}\frac{\Delta z}{\Delta x} = $$

$$= \frac{f(z+\Delta z)-f(z)}{z}k.$$

Сделаем предельный переход. Пусть $\Delta x \to 0$. Тогда, по принципу непрерывности, $\Delta z \to 0$,

$$\frac{f(z+\Delta z)-f(z)}{z} \to f'(z);\ k\frac{f(z+\Delta z)-f(z)}{z} \to kf'(z).$$

Итак, $\frac{\Delta g}{\Delta x} \to kf'(z) = kf'(kx + b)$, что и требовалось доказать.

Примеры

1. $y = (x + 1)^5$, $y' = 5(x + 1)^4$.
2. $y = (5x - 1)^4$, $y' = 5 \cdot 4\,(5x - 1)^3 = 20\,(5x - 1)^3$.
3. $y = \frac{1}{2x+3}$, $y' = -2\frac{1}{(2x+3)^2} = -\frac{2}{(2x+3)^2}$.
4. $y = \sqrt{6x-1}$, $y' = 6\ \frac{1}{2\sqrt{6x-1}} = \frac{3}{\sqrt{6x-1}}$.

При дифференцировании функций вида $y = (kx)^n$ выведенной формулой пользоваться нерационально. Лучше вынести константу за знак производной.

Пример

$$((3x)^5)' = 3^5 \cdot (x^5)' = 3^5 \cdot 5 \cdot x^4.$$

Координаты вектора скорости

Если положение точки задается вектором r = r(t), имеющим координаты $x = x(t)$ и $y = y(t)$, то вектор скорости v = v(t) имеет координаты $x'(t)$, $y'(t)$, равные производным координат вектора r:

$$\mathbf{r}'(x; y) = \mathbf{v}(x'; y').$$

Пусть точка A движется по криволинейной траектории. Обозначим координаты точки A в момент времени t через $x(t)$ и $y(t)$ (рис. 316). Эти координаты зависят от времени t и являются функциями от t. Рассмотрим мгновенную скорость v движущейся точки A в момент времени t. Вектор мгновенной скорости v направлен по касательной к траектории в точке A. Координаты вектора v тоже зависят от времени t и меняются, вообще говоря, при переходе от одной точки траектории к другой.

Покажем, что координаты вектора скорости v точки A равны $x'(t)$ и $y'(t)$, где $x'(t)$ и $y'(t)$ — производные функций x и y в точке t.

За время Δt точка A переместится в точку B с координатами $x(t + \Delta t)$ и $y(t + \Delta t)$ (рис. 317). Рассмотрим вектор перемещения $\overline{AB}$. Он является разностью векторов $\overline{OB}$ и $\overline{OA}$, и его координаты поэтому будут разностями координат этих векторов. Следовательно, координаты вектора $\overline{AB}$

$$(x(t + \Delta t) - x(t); y(t + \Delta t) - y(t)).$$

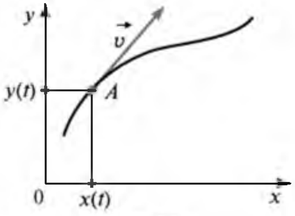


Рис. 316

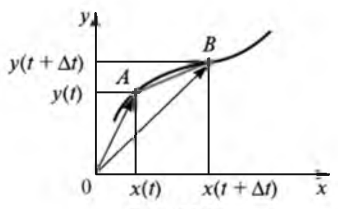


Рис. 317

Напомним, что *средней скоростью* называется отношение перемещения ко времени. Таким образом, вектор средней скорости равен $\dfrac{\overline{AB}}{\Delta t}$, поэтому его координаты

$$\left(\frac{x(t+\Delta t)-x(t)}{\Delta t}; \frac{y(t+\Delta t)-y(t)}{\Delta t}\right).$$

Вектор средней скорости при уменьшении Δt приближается к вектору мгновенной скорости в момент времени t. (Это значит, что разность между вектором мгновенной скорости и средней скоростью станет сколь угодно малой при $\Delta t \to 0$.)

По определению производной, при уменьшении Δt величина $\dfrac{x(t+\Delta t)-x(t)}{t}$ приближается к $x'(t)$, а $\dfrac{y(t+\Delta t)-y(t)}{\Delta t}$ – к $y'(t)$. Таким образом,

координаты вектора мгновенной скорости в момент времени t равны $x'(t)$ и $y'(t)$.

Пример

Рассмотрим снова движение снаряда, выпущенного под углом α к горизонту с начальной скоростью $\mathbf{v}_0$ (рис. 318). Снаряд, если пренебречь сопротивлением воздуха, будет двигаться по параболе. Вектор скорости $\mathbf{v}$ направлен по касательной к ней. Известное соотношение $\mathbf{v} = \mathbf{v}_0 + \mathbf{g}t$ получается дифференцированием соотношения $\mathbf{r} = \mathbf{v}_0 t + \dfrac{1}{2}\mathbf{g}t^2$ (векторы $\mathbf{v}_0$ и $\mathbf{g}$ постоянны). Можно вычислить и координаты вектора скорости, зная координаты вектора $\mathbf{r}(x; y)$:

$$x(t) = v_0 t\cos\alpha;$$

$$y(t) = v_0 t\sin\alpha - \frac{1}{2}gt^2.$$

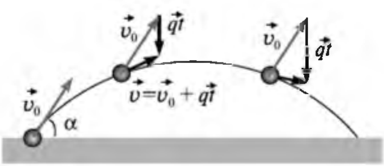


Рис. 318

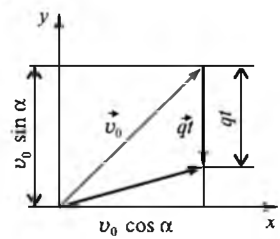


Рис. 319

Известно, что вектор скорости $\mathbf{v}$ имеет координаты $(x'(t), y'(t))$. Вычислим производные: $x' = v_0\cos\alpha$; $y' = v_0\sin\alpha - gt$. Эти же равенства можно получить, спроектировав известную заранее векторную формулу $\mathbf{v} = \mathbf{v}_0 + \mathbf{g}t$ на оси координат (рис. 319).

Производные тригонометрических функций

Производные синуса и косинуса

Пусть точка A движется с единичной скоростью по окружности радиуса 1 с центром в начале координат в положительном направлении. Координаты точки $A = P_t$ в момент времени t равны $\cos t$ и $\sin t$.

Вектор мгновенной скорости точки A в момент времени t направлен по касательной к окружности в точке A (рис. 320), и, по теореме о перпендикулярности касательной радиусу, проведенному в точку касания, вектор **v** перпендикулярен радиусу-вектору $\overrightarrow{OA}$.

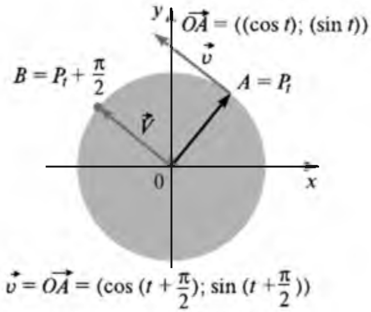


Рис. 320

Вычислим координаты вектора **v**. Отложив от точки O вектор **v**, мы получим вектор $\overrightarrow{OB}$, координаты которого равны координатам вектора **v**. Так как движение точки A по окружности происходит с единичной скоростью, то длина вектора равна 1 и поэтому длина вектора $\overrightarrow{OB}$ также равна 1. Следовательно, точка B лежит на окружности.

Вектор $\overrightarrow{OB}$ перпендикулярен вектору $\overrightarrow{OA}$, поэтому если $A = P_t$, то $B = P_t + \frac{\pi}{2}$. Таким образом, координаты вектора $\mathbf{v} = \overrightarrow{OB}$ равны

$$\cos\left(t + \frac{\pi}{2}\right) = -\sin t \quad \text{и} \quad \sin\left(t + \frac{\pi}{2}\right) = \cos t.$$

Найденные значения координат вектора **v** являются производными от координат радиуса-вектора точки A, равных $\sin t$, $\cos t$. Следовательно,

$$(\sin t)' = \cos t,\ (\cos t)' = -\sin t.$$

Перейдя к обозначениям $\sin x$, $\cos x$, получим табл. 4 производных синуса и косинуса.

Таблица 4

Производные синуса и косинуса

y	y'
$\sin x$	$\cos x$
$\cos x$	$-\sin x$

Примеры

Найти производную функции:

а) $y = A \sin (\omega t + \alpha)$;

$$y' = A \cos (\omega t + \alpha)\, \omega = A\omega \cos (\omega t + \alpha).$$

Производная содержит множитель ω (скорость колебания пропорциональна его частоте).

б) $y = \sin x \cos x$;

$$y' = (\sin x)' \cos x + \sin x\,(\cos x)' = \cos^2 x - \sin^2 x;$$

в) $y = \sin 5x$;

$$y' = 5\cos 5x;$$

г) $y = \cos\dfrac{x}{3}$;

$$y' = -\frac{1}{3}\sin\frac{x}{3};$$

д) $y = 2\sin\left(3x - \dfrac{\pi}{6}\right)$;

$$y' = 2 \cdot 3 \cdot \cos\left(3x - \frac{\pi}{6}\right) = 6\cos\left(3x - \frac{\pi}{6}\right);$$

е) $y = 5\cos\left(\dfrac{\pi}{3} - x\right)$;

$$y' = 5 \cdot (-1) \cdot (-1)\sin\left(\frac{\pi}{3} - x\right) = 5\sin\left(\frac{\pi}{3} - x\right).$$

Производные тангенса и котангенса

Вычислим теперь производную функции $y = \operatorname{tg} x$. Так как $\operatorname{tg} x = \dfrac{\sin x}{\cos x}$, то по теореме о производной частного получаем

$$(\operatorname{tg} x)' = \frac{(\sin x)'\cos x - (\cos x)'\sin x}{\cos^2 x} = \frac{\cos^2 x + \sin^2 x}{\cos^2 x} = \frac{1}{\cos^2 x}.$$

Следовательно,

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

Аналогично,

$$(\operatorname{ctg} x)' = \left(\frac{\cos x}{\sin x}\right)' = \frac{(\cos x)'\sin x - (\sin x)'\cos x}{\sin^2 x} = \frac{-\sin^2 x - \cos^2 x}{\sin^2 x} = -\frac{1}{\sin^2 x}.$$

Таким образом,

$$(\operatorname{ctg} x)' = -\frac{1}{\sin^2 x}.$$

Полученные значения производных тангенса и котангенса сведем в табл. 5:

Таблица 5

Производные тангенса и котангенса

y	y'
$\operatorname{tg} x$	$\dfrac{1}{\cos^2 x}$
$\operatorname{ctg} x$	$-\dfrac{1}{\sin^2 x}$

Примеры

1. $y = \operatorname{tg} x + \operatorname{ctg} x$.

$$y' = \frac{1}{\cos^2 x} - \frac{1}{\sin^2 x} = \frac{\sin^2 x - \cos^2 x}{\sin^2 x \cos^2 x}.$$

2. $y = 2 \operatorname{tg} 3x$.

$$y' = 2 \cdot 3 \frac{1}{\cos^2 3x} = \frac{6}{\cos^2 3x}.$$

3. $y = 3 \operatorname{ctg} \left(\frac{x}{2} - \frac{\pi}{4}\right)$.

$$y' = -3 \cdot \ y' = -3 \cdot \frac{1}{2} \cdot \left(\frac{1}{\sin^2(x/2 - \pi/4)}\right)^2 = -\frac{3}{2\sin(x/2 - \pi/4)}.$$

Дифференциал

Основой разнообразных физических приложении производной является понятие дифференциала.

Дифференциалом функции называют произведение ее производной на приращение аргумента.

Пусть задана функция $y = f(x)$. Ее дифференциал обозначают dy (или можно писать df). По определению,

$$dy = f'(x)\Delta x.$$

Рассмотрим функцию $y = x$. Так как производная этой функции постоянна и равна единице, то дифференциал этой функции равен

Δx, т. е. для независимой переменной верно равенство $\mathrm{d}x = \Delta x$. Поэтому дифференциал функции записывают обычно в виде $\mathrm{d}y = f'(x)\mathrm{d}x$. Из этой записи происходит одно из обозначений производной: $f'(x) = \dfrac{\mathrm{d}y}{\mathrm{d}x}$, которое можно понимать как отношение дифференциалов

Какими свойствами обладает введенное нами понятие дифференциала функции?

Дифференциал функции — это главная часть ее приращения.

Сравним приращение функции Δy и ее дифференциал dy (рис. 321, *а*). На рисунке видно, что чем меньше Δx, тем ближе Δy к $\mathrm{d}y$. Действительно, разность $\Delta y - \mathrm{d}y$ преобразуем так:

$$\Delta y - \mathrm{d}y = \Delta y - f'(x_0)\Delta x = \Delta x\left(\frac{\Delta y}{\Delta x} - f'(x_0)\right).$$

По определению производной, при стремлении Δx к нулю разность $\dfrac{\Delta y}{\Delta x} - f'(x_0)$ тоже стремится к нулю. При умножении на Δx мы получим выражение, еще быстрее приближающееся к нулю. Таким образом, приращение функции отличается от ее дифференциала на такую функцию, которая стремится к нулю еще быстрее, чем Δx. Поэтому и говорят, что дифференциал есть главная часть приращения функции.

Предупреждаю, чтобы остерегались отбрасывать d*x* — ошибка, которую часто допускают и которая препятствует продвижению вперед.
Г. В. Лейбниц

Дифференциал функции — это линейная функция от приращения аргумента.

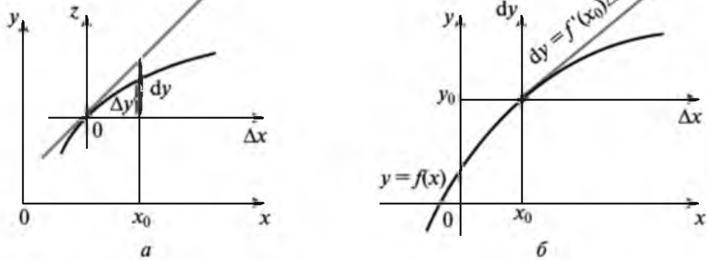


Рис. 321

Если мы зафиксируем точку на графике функции (x_0, y_0) и обозначим $f'(x_0)$) через k, то на равенство $\mathrm{d}y = k\mathrm{d}x$ можно смотреть как на линейную зависимость между $\mathrm{d}y$ и $\mathrm{d}x$. Если через точку (x_0, y_0) провести новые оси координат (их можно обозначить $\mathrm{d}x$ и $\mathrm{d}y$, то график этой зависимости будет прямой, касательной к графику функции f в точке (x_0, y_0) (рис. 321, *б*). Можно сказать, что дифференциал функции f — это линейная функция, графиком которой является касательная к графику f Геометрически на соотношение $\mathrm{d}y = f'(x_0)\mathrm{d}x$ можно смотреть как на уравнение касательной к графику функции f, записанное в системе координат $(\mathrm{d}x, \mathrm{d}y)$. Данная система координат получается из исходной параллельным переносом начала координат в точку (x_0, y_0). Это позволяет получить

уравнение касательной в исходных координатах:

$$y - y_0 = f'(x_0)(x - x_0).$$

Вывод уравнения касательной легко запомнить. Прежде всего касательная проходит через заданную точку (x_0, y_0) на графике функции $y = f(x)$. Поэтому ее уравнение $y - y_0 = k(x - x_0)$. Угловой коэффициент равен значению производной в заданной точке:

$$k = f'(x_0).$$

Дифференциал в физике

Мы ввели понятие дифференциала с помощью равенства $\mathrm{d}y = f'(x)\mathrm{d}x$. Для вычисления дифференциала надо найти производную. Однако, помня о том, что дифференциал — это главная часть приращения функции, линейно зависящая от приращения аргумента, мы из физических соображений получим равенство вида $\mathrm{d}y = k\mathrm{d}x$ и сделаем вывод о том, что k — это производная y по x.

1. *Работа*. Найдем работу, которую совершает заданная сила F при перемещении по отрезку оси x. Если сила F постоянна, то работа A равна произведению F на длину пути. Если сила меняется, то ее можно рассматривать как функцию от x: $F = F(x)$. Приращение работы ΔA на отрезке $[x, x + \mathrm{d}x]$ нельзя точно вычислить как произведение $F(x)\mathrm{d}x$, так как сила меняется на этом отрезке. Однако при маленьких $\mathrm{d}x$ можно считать, что сила меняется незначительно и произведение представляет главную часть ΔA, т. е. является дифференциалом работы ($\mathrm{d}A = F(x)\mathrm{d}x$). Таким образом, силу можно считать призводной работы по перемещению.

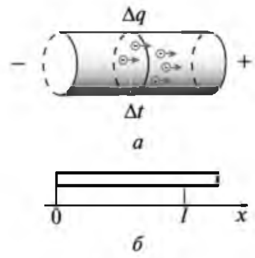


Рис. 322

2. *Заряд.* Пусть q — заряд, переносимый электрическим током через поперечное сечение проводника за время t. Если сила тока I постоянна, то за время $\mathrm{d}t$ ток перенесет заряд, равный $I\mathrm{d}t$. При силе тока, изменяющейся со временем по закону $I = I(t)$, произведение $I(t)\mathrm{d}t$ дает главную часть приращения заряда на маленьком отрезке времени $[t, t + \mathrm{d}t]$, т. е. является дифференциалом заряда: $\mathrm{d}q = I(t)\mathrm{d}t$. Следовательно, сила тока является производной заряда по времени (рис. 322, *а*).

3. *Масса тонкого стержня.* Пусть имеется неоднородный тонкий стержень. Если ввести координаты так, как показано на рис. 322, *б*, то функция $m = m(l)$ — масса куска стержня от точки O до точки l. Неоднородность стержня означает, что его линейная плотность не является постоянной, а зависит от положения точки l по некоторому закону $\rho - \rho(l)$. Если на маленьком отрезке стержня $[l, l + \mathrm{d}l]$ предположить, что плотность постоянна и равна $\rho(l)$, то произведение $\rho(l)\mathrm{d}l$ дает дифференциал массы $\mathrm{d}m$. Значит, линейная плотность — это производная массы по длине.

4. *Теплота.* Рассмотрим процесс нагревания какого-нибудь вещества и вычислим количество теплоты $Q(T)$, которое необходимо, чтобы нагреть 1 кг вещества от 0 °С до T. Зависимость $Q = Q(T)$ очень сложна и определяется экспериментально. Если бы теплоемкость c данного вещества не зависела от температуры, то произведение $c\mathrm{d}T$ дало бы изменение количества теплоты. Считая на малом отрезке $[T, T + \mathrm{d}T]$ теплоемкость постоянной, получаем дифференциал количества теплоты $\mathrm{d}Q = c(T)\mathrm{d}T$. Поэтому теплоемкость — это производная теплоты по температуре.

5. *Снова работа.* Рассмотрим работу как функцию времени. Нам известна характеристика работы, определяющая ее скорость по времени, — это мощность. При работе с постоянной мощностью N. работа за время $\mathrm{d}t$ равна $N\mathrm{d}t$. Это выражение представляет дифференциал работы, т. е. $\mathrm{d}A = N(t)\mathrm{d}t$, и мощность выступает как производная работы по времени.

Все приведенные примеры были построены по одному и тому же образцу. В каждом примере шла речь о связи между тремя величинами, знакомыми нам из курса физики: работа, перемещение, сила; заряд, время, сила тока; масса, длина, линейная плотность; и т. д. Каждый раз одна из этих величин выступала как коэффициент пропорциональности между дифференциалами двух других, т. е. каждый раз появлялось соотношение вида $dy = k(x)dx$. На такое соотношение можно смотреть как на способ определения величины $k(x)$. Тогда $k(x)$ находится (или определяется) как производная y по x. Этот вывод мы и фиксировали в каждом примере. Возможна и обратная постановка вопроса; как найти зависимость y от x из заданного соотношения между их дифференциалами. Эту задачу мы рассмотрим в главе, посвященной интегрированию.

Приближенные формулы

Главная приближенная формула вблизи нуля $\sin t \approx t$.

Доказательство. Найдем дифференциал функции $y = \sin x$ при $x = 0$: $dy = \cos x\, dx$. Так как $\cos 0 = 1$, то при $x = 0$ $d \sin x = dx$. Так как $\Delta y \approx dy$, то $\Delta y = \sin \Delta x \approx dy = dx = \Delta x$. Вместо Δx возьмем t и получим $\sin t \approx t$. Необходимо помнить, что во всех этих приближенных формулах аргумент выбирается близким к нулю.

Эта формула дает тем точнее значение синуса, чем ближе t к нулю. Возможность заменять $\sin t$ на t при малых значениях угла t широко используется в приближенных вычислениях. Можно дать различные интерпретации приближенной формулы $\sin t \approx t$.

1. $\lim\limits_{t\to 0}\dfrac{\sin t}{t}=1$ — отношение приращения функции и его главной части стремится к единице при стремлении к нулю приращения аргумента.
2. Рассмотрим единичный круг. Пусть для простоты $t > 0$. Тогда длина дуги AB равна t, а длина отрезка BC равна $\sin t$. Удобно удвоить дуги AB и отрезок BC — дуга BD имеет длину $2t$, а хорда BD — $2 \sin t$ (рис. 323). Соотношение $\sin t \approx t$ означает, что отношение длины хорды к длине стягиваемой ею дуги стремится к единице, когда дуга стягивается в точку.
3. Рассмотрим касательную к синусоиде в начале координат.

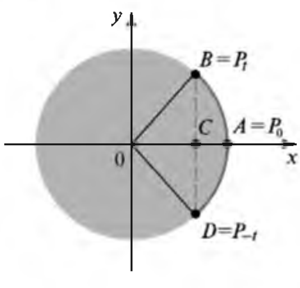


Рис. 323

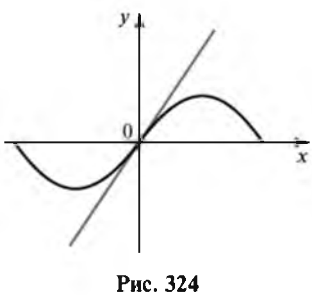


Рис. 324

Так как $(\sin x)' = \cos x$, а $\cos 0 = 1$, то уравнение этой касательной $y = x$ (рис. 324). Таким образом, заменяя вблизи начала координат график синуса отрезком касательной, мы вычисляем приближенное значение синуса по формуле $\sin x \approx x$.

Для получения других приближенных формул выпишем дифференциалы тангенса и косинуса:

$$y = \operatorname{tg} x,\ \mathrm{d}y = \frac{1}{\cos^2 x}\,\mathrm{d}x;$$
$$y = \cos x,\ \mathrm{d}y = -\sin x\,\mathrm{d}x.$$

При $x = 0$ получим приближенное значение тангенса:

$$\operatorname{tg} \Delta x \approx \Delta x, \text{ т.е. } \operatorname{tg} t \approx t.$$

Применяя этот же прием к косинусу, находим, что дифференциал косинуса при $x_0 = 0$ равен $-\sin 0\,\mathrm{d}x = 0$. Это означает, что главная часть приращения косинуса равна нулю и в первом приближении $\cos x \approx \cos 0 = 1$. Можно получить более точную формулу таким путем. Запишем $\cos x$ так: $\cos x = \sqrt{1-\sin^2 x}$. Заменим в этой формуле $\sin x$ на x и воспользуемся приближенной формулой для корня: $\cos x = \sqrt{1-\sin^2 x} \approx \sqrt{1-x^2} \approx 1 - \frac{1}{2}x^2$. Эта формула для косинуса вблизи точки $x_0 = 0$ весьма точна. Более точные приближения можно получить с помощью формулы Тейлора. Для синуса и косинуса формула Тейлора имеет вид

$$\sin x = x - \frac{x^3}{6} + \frac{x^5}{120} - \ldots,\quad \cos x = 1 - \frac{x^2}{2} + \frac{x^4}{24} - \ldots$$

Примеры

Вычислить приближенно:

а) $\sin 0{,}03 \cdot \operatorname{tg} 0{,}12$; $\sin 0{,}03 \approx 0{,}03$; $\operatorname{tg} 0{,}12 \approx 0{,}12$; $\sin 0{,}03 \cdot \operatorname{tg} 0{,}12 \approx \approx 0{,}0036 \approx 0{,}004$;

б) $\sin 2^\circ$. Переводим 2° в радианную меру: $2^\circ \approx 0{,}034$, поэтому $\sin 2^\circ \approx 0{,}034$;

в) $\sin 32^\circ$. Выберем исходное «хорошее» значение аргумента: $x_0 = 30^\circ$. Представляем $x = 32^\circ = 30^\circ + 2^\circ = x + \mathrm{d}x$. При подстановке заменяем в $\mathrm{d}x$ градусную меру на радианную: $\sin 32^\circ \approx 0{,}5 + \frac{\sqrt{3}}{2 \cdot 0{,}034} \approx 0{,}529$.

Производная показательной функции

Возьмем показательную функцию $y = a^x$. Вычислим среднюю скорость роста этой функции на отрезке $[x, x + \Delta x]$:

$$\frac{\Delta y}{\Delta x} = \frac{a^{x+\Delta x} - a^x}{\Delta x} = a^x \frac{a^{\Delta x} - 1}{\Delta x}.$$

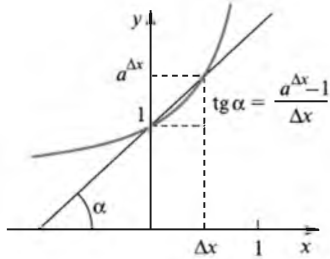


Рис. 325

Мы видим, что средняя скорость роста показательной функции в точке x равна значению этой функции в точке x, умноженному на число $\frac{a^{\Delta x} - 1}{\Delta x}$. Исследуем, как ведет себя эта дробь при малых значениях Δx. Так как $a^0 = 1$, то $a^{\Delta x}$ при малых значениях Δx близко к 1. Если провести секущую на графике функции, проходящую через точки (0; 1) и (Δx; $a^{\Delta x}$), то ее угловой коэффициент будет равен $\frac{a^{\Delta x} - 1}{\Delta x}$ (рис. 325). При $\Delta x \to 0$ секущая приближается к касательной к графику функции в точке (0; 1). Это означает, что $\frac{\Delta y}{\Delta x}$ приближается к произведению a^x на значение производной при $x = 0$. Итак,

---

для нахождения производной функции $y = ax$ надо знать значение этой производной в нуле.

Обозначим ее через k, тогда

$$(a^x)' = ka^x,$$

т.е. производная показательной функции пропорциональна самой функции.

Как же найти коэффициент пропорциональности k? Мы знаем, что он равен угловому коэффициенту касательной, проведенной в точке (0; 1). Можно приближенно по графику вычислить этот коэффициент. Так, известно, что для $a = 10$ его значение $k \approx 2{,}3$, поэтому $(10^x)' \approx 2{,}3 \cdot 10^x$.

Число e.

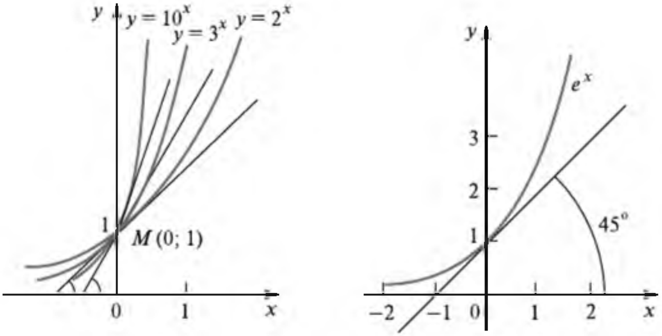
Посмотрим на графики показательных функций при различных $a > 1$ (рис. 326, *а*). Все они проходят через точку M(0; 1). Проведем в этой точке касательные к графикам. Чем больше основание a, тем «круче» расположена касательная. Так, при $a = 2$ угловой коэффициент касательной равен 0,693, а при $a = 10$ равен 2,303. Ясно, что при непрерывном изменении a от 2 до 10 угловой коэффициент касательной в точке M непрерывно изменяется и найдется такое значение a, для которого он равен 1. Такое основание a обозначается буквой e. Число e иррационально. Его приближенное значение таково: $e \approx 2{,}718$. Итак,

число *e* — это такое основание, при котором угловой коэффициент касательной к графику функции $y = ex$ в точке $x = 0$ равен единице, т.е. касательная в этой точке образует с осью абсцисс угол 45° (рис. 326, *б*).

Это свойство можно сформулировать иначе. Производная любой показательной функции пропорциональна самой этой функции. Число *e* — это основание, для которого коэффициент пропорциональности равен единице, т.е.

***e* — это такое число, что производная функции $y = ex$ равна самой этой функции:**

$$(e^x)' = e^x.$$



а *б*

Рис. 326

Функцию $y = e^x$ часто обозначают $y = \exp x$ (читается: «эксп от *x*») и называют *экспонентой*. Экспонентами называют и функции более общего вида: $y = ce^{kx}$. Подумайте, понятен ли вам смысл таких распространенных выражений: «Численность бактерий растет по экспоненте», «Сила тока затухает по экспоненте», «Его успехи растут по экспоненте».

С помощью числа *e* легко записать коэффициент пропорциональности *k* для любого основания *a*. Для этого заменим основание *a* на степень основания *e*: $a^x = \left(e^{\log_e a}\right)^x = e^{(\log_e a)x}$. Тогда $(a^x)' = \left(e^{(\log_e a)x}\right)' =$ $= e^{(\log_e a)x} \log_e a = \log_e a \cdot a^x$.

С помощью символа lim коэффициент *k* можно записать следующим образом:

$$k = \lim_{\Delta x \to 0} \frac{a^{\Delta x} - 1}{\Delta x}.$$

Мы выяснили, что число k есть логарифм a по некоторому основанию e. Для $a = e$ число $k = 1$, тогда

$$\lim_{\Delta x \to 0} \frac{e^{\Delta x} - 1}{\Delta x} = 1.$$

Это предельное соотношение часто служит определением числа e. Пользуясь приближенными равенствами, находим

$$\frac{e^{\Delta x} - 1}{\Delta x} \approx 1 \Rightarrow e^{\Delta x} - 1 \approx \Delta x \Rightarrow$$

$$\Rightarrow e^{\Delta x} \approx 1 + \Delta x \Rightarrow e \approx (1 + \Delta x)^{\frac{1}{\Delta x}}, \text{ т.е.}$$

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

Если в качестве стандартной переменной Δx, стремящейся к нулю, взять последовательность $\Delta x = \frac{1}{n}$ $(n = 1, 2, 3, \ldots)$, то

$$e = \lim_{n \to \infty} \left(1 + \frac{1}{n}\right)^n.$$

Число e играет исключительно важную роль. При нахождении производной произвольной показательной функции мы пришли к необходимости рассматривать логарифмы по основанию e. Эти логарифмы получили название натуральных. Натуральные логарифмы были введены шотландским математиком Дж. Непером в начале XVII в., фактически раньше, чем логарифмы по другим основаниям. Натуральный логарифм обозначается ln.

Теперь можно записать:

$$a^x = e^{x \ln a}, \quad (a^x)' = \ln a \cdot a^x.$$

При переходе от десятичных логарифмов к натуральным можно пользоваться модулем перехода:

$$\lg x = \frac{\ln x}{\ln 10} = \frac{\ln x}{2{,}303} = 0{,}434 \ln x,$$

$$\ln x = \frac{\lg x}{\lg e} = \ln 10 \lg x = 2{,}303 \lg x.$$

Производная логарифмической функции

Рассмотрим две функции: $y = e^x$ и $y = \ln x$. Мы знаем, что их графики симметричны относительно прямой $y = x$. Это поможет нам найти производную логарифмической функции, зная производную экспоненты. Возьмем точку $P(c; d)$ на графике экспоненты (т.е. $d = e^c$) и симметричную точку $Q(d; c)$ на графике логарифмической функции (т.е. $c = \ln d$). Касательные к графикам в этих точках тоже будут симметричны (рис. 327, *а*). Угловой коэффициент k касательной к графику экспоненты равен значению производной функции $y = e^x$ при $x = c$, т.е. $k_1 = e^c$, так как $(e^x)' = e^x$. Пусть α_1 и α_2 — углы, образованные проведенными касательными с осью абсцисс. Из рисунка 327, *б* ясно, что $\alpha_1 + \alpha_2 = \frac{\pi}{2}$. Так как $k_1 = \operatorname{tg}\alpha_1 = e^c$, то $k_2 = \operatorname{tg}\alpha_2 = \operatorname{tg}\left(\frac{\pi}{2} - \alpha_1\right) = \operatorname{ctg}\alpha_1 = \frac{1}{k_1} = \frac{1}{e^c} = \frac{1}{d}$. Таким образом, производная функция $y = \ln x$ в точке $x = d$ равна $\frac{1}{d}$. Тогда

$$(\ln x)' = \frac{1}{x}.$$

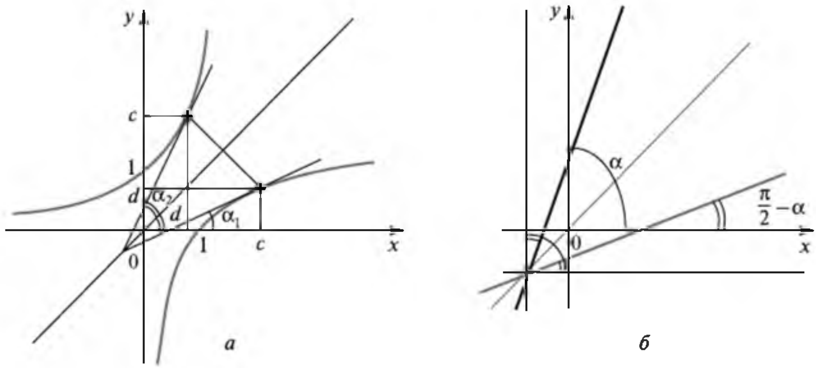


Рис. 327

Мы видим, что производная логарифмической функции $y = \ln x$ равна степенной функции $y = x^{-1}$. Интересно заметить, что функция $y = x^{-1}$ не получается как производная какой-либо другой степенной функции вида $y = cx^k$. Действительно, хотя $(x^k)' = kx^{k-1}$ при любом k, но получить значение $k - 1$, равное -1, можно лишь при $k = 0$, а $(x^0)' = 0$.

Так как $\log_a x = \frac{1}{\ln a}\ln x$, то $(\log_a x)' = \frac{1}{\ln a}\frac{1}{x}$. По формулам производной показательной функции $(e^x)' = e^x$ и $(a^x)' = ka^x$, где $k = \lim\limits_{\Delta x \to 0}\frac{a^{\Delta x} - 1}{\Delta x}$, известно, что $e^{\ln a} = a$ и $a^x = e^{kx}$, где $k = \ln a$. Поэтому $(a^x)' = ke^{kx} = ka^x$, т.е. $(a^x)' = (\ln a)\, a^x$.

Зная производные экспоненты и логарифма, можно получить приближенные формулы для их вычисления.

Пусть $y = e^x$, $x_0 = 0$, $y_0 = e^0 = 1$. Разность $e^h - e^0 = e^h - 1$ — это приращение y на отрезке $[0; h]$. Вычислив $\mathrm{d}y$ при $x_0 = 0$, получим $\mathrm{d}y = y'(0)\,\mathrm{d}x$. Так как $y' = e^x$, то $y'(0) = 1$. Заменив Δy на $\mathrm{d}y$ и подставив $\mathrm{d}x = h$, получим приближенную формулу

$$e^h - 1 \approx h \text{ или } e^h \approx 1 + h.$$

Более точная формула для вычисления экспоненты такова:

$$e^h = 1 + h + \frac{h^2}{2} + \frac{h^3}{6} + \frac{h^4}{24} + \ldots$$

Пусть теперь $y = \ln x$. Выберем $x_0 = 1$, $y_0 = \ln 1 = 0$. Положим $\mathrm{d}x = h$ и вычислим $\ln(1+h)$. Найдем $\mathrm{d}y$ при $x_0 = 1$. Так как $(\ln x)' = \frac{1}{x}$, то $y'(x_0) = 1$ и $\mathrm{d}y = 1 \cdot \mathrm{d}x = h$. Заменяя $\Delta y = \ln(1+h) - \ln 1 = \ln(1+h)$, получаем приближенную формулу

$$\ln(1+h) \approx h.$$

Более точная формула для вычисления логарифма такова:

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

Примеры

1. $y = xe^x$; $y' = e^x + xe^x = e^x(1+x)$.
2. $y = x\ln x - x$; $y' = \ln x + \frac{x}{x} - 1 = \ln x$.
3. $y = e^{kx}$; $y' = ke^{kx}$.
4. $y = \ln kx$; $y' = \frac{k}{kx} = \frac{1}{x}$ или $\ln kx = \ln k + \ln x$; $(\ln kx)' = \frac{1}{x}$.

Производная обратной функции

Выведем общую формулу производной обратной функции аналогично тому, как мы выводили формулу производной логарифмической функции. Пусть f и g — взаимно обратные функции. Как найти производную функции g, зная производную функции f?

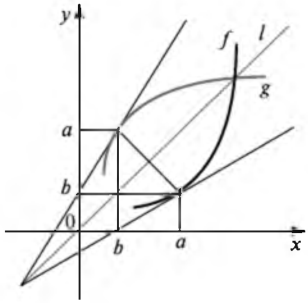


Рис. 328

Графики функций $y = f(x)$ и $y = g(x)$ симметричны друг другу относительно биссектрисы l угла xOy (рис. 328).

Возьмем какую-нибудь точку $x = a$ и вычислим значение одной из функций в этой точке: $f(a) = b$. Тогда, по определению обратной функции, $g(b) = a$.

Точки $(a; f(a)) = (a; b)$ и $(b; g(b)) = (b; a)$ симметричны относительно указанной прямой l. Так как кривые симметричны, то и касательные к ним симметричны относительно прямой l.

Из симметрии ясно, что угол одной из этих прямых с осью x равен углу другой прямой с осью y. Если первая прямая образует с осью x угол α, то ее угловой коэффициент равен $k_1 = \operatorname{tg}\alpha$, тогда вторая прямая имеет угловой коэффициент $k_2 = \operatorname{tg}\left(\frac{\pi}{2} - \alpha\right) = \operatorname{ctg}\alpha = \frac{1}{k_1}$. Таким образом, угловые коэффициенты прямых, симметричных относительно прямой l, взаимно обратны, т.е. $k_2 = \frac{1}{k_1}$ или $k_1 k_2 = 1$.

Переходя к производным и учитывая, что угловой коэффициент касательной является значением производной в точке касания, делаем вывод:

значения производных взаимно обратных функций в соответствующих точках взаимно обратны:

$$g'(b) = \frac{1}{f'(a)} = \frac{1}{f'(g(b))}.$$

Напомним еще, что $b = f(a)$ и $a = g(b)$.

В приведенных рассуждениях предполагалось, что $k_1 \neq 0$, т.е. касательные к кривым не параллельны осям координат.

Примеры

1. $y = f(x) = x^3$. Обратной функцией будет функция $y = g(x) = \sqrt[3]{x}$. Найдем производную функции g:

$$g'(x) = \frac{1}{f'(g(x))} = \frac{1}{3(g(x))^2} = \frac{1}{3\sqrt[3]{x^2}}, \text{ т.е. } (\sqrt[3]{x})' = \frac{1}{3\sqrt[3]{x^2}}.$$

2. $y = f(x) = \sin x, x \in \left[-\frac{\pi}{2}; \frac{\pi}{2}\right]$.

   Обратной функцией будет $y = g(x) = \arcsin x$. Найдем производную арксинуса:

$$g'(x) = \frac{1}{f'(g(x))} = \frac{1}{\cos \arcsin x} = \frac{1}{\sqrt{1-x^2}}.$$

   Итак, $(\arcsin x)' = \dfrac{1}{\sqrt{1-x^2}}$.

   Аналогично вычисляется производная арктангенса:

$$(\operatorname{arctg} x)' = \frac{1}{\dfrac{1}{\cos^2 \operatorname{arctg} x}} = \cos^2 \operatorname{arctg} x = \frac{1}{1+\operatorname{tg}^2 \operatorname{arctg} x} = \frac{1}{1+x^2}.$$

Контрольные вопросы и задания

1. Какова схема вычисления производной?
2. Чему равна производная суммы? Производная произведения? Производная частного двух функций?
3. Что происходит с производной при умножении функции на некоторую постоянную?
4. Чему равна производная степенной функции?
5. Как меняется производная при линейной замене аргумента?
6. Чему равна производная синуса, косинуса, тангенса, котангенса?
7. Каковы координаты вектора скорости вращательного движения?
8. По каким приближенным формулам можно вычислить синус, косинус и тангенс угла, близкого к нулю?
9. Каково основное свойство производной показательной функции?
10. Известно, что производная показательной функции пропорциональна самой функции. Как найти коэффициент пропорциональности?
11. Как определяется число e?
12. Чему приближенно равно число e?
13. Чему равна производная функции $y = e^x$?
14. Через какую точку проходят графики всех показательных функций $y = a^x$?

ГЛАВА 8

ИЗМЕРЕНИЯ

§ 32. Определение интеграла

Задача интегрирования

Математика изучает различные связи между величинами. Важнейшие примеры таких связей дает механическое движение. Мы уже много раз обращались к примеру движения материальной точки на оси. Между положением (координатой) точки и ее скоростью есть известная связь, лежащая в основе математического анализа: скорость является производной от координаты по времени. Операция нахождения производной называется дифференцированием. Обратная задача — нахождение положения точки по ее скорости — решается с помощью другой математической операции — интегрирования.

Мы знаем много пар величин, которые связаны между собой так же, как положение точки и ее скорость. Нахождение одной из этих величин, если известна другая, сводится к операции дифференцирования. Так, линейная плотность тонкого стержня есть производная его массы по длине, мощность есть производная работы по времени, сила тока есть производная заряда по времени и др. С помощью обратной операции интегрирования, мы научимся вычислять массу по заданной плотности, работу по известной мощности, заряд по заданной силе тока и др.

Прежде чем учиться вычислять интегралы, рассмотрим их геометрический смысл. Начнем по-прежнему с задачи о механическом движении. Пусть точка движется с постоянной скоростью $v = v_0$. Графиком скорости в системе координат $(t; v)$ будет прямая $v = v_0$, параллельная оси времени t (рис. 329). Если считать, что в начальный момент времени $(t = 0)$ точка находилась в начале координат, то ее

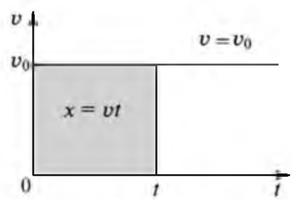


Рис. 329

положение x в момент времени t запишется формулой $x = vt$. Величина vt представляет собой площадь прямоугольника, ограниченного графиком скорости, осью абсцисс и двумя вертикальными прямыми, т.е. путь (положение точки) можно вычислить как площадь под графиком скорости. Аналогичная связь между величинами сохраняется и при движении точки с непостоянной скоростью. Так, мы знаем, что в случае движения с ускорением a и начальной скоростью $v_0 = 0$ скорость изменяется по закону $v = at$, а положение точки (считая, что она находилась при $t = 0$ в начале координат) — по закону $x = \frac{at^2}{2}$. График скорости — прямая, а площадь под ней вычисляется по формуле площади треугольника: $S = \frac{at \cdot t}{2} = \frac{at^2}{2}$, т.е., как и прежде, путь можно найти, как площадь под графиком скорости.

Таким образом, задача интегрирования тесно связана с задачей вычисления площади.

Геометрический смысл интеграла

Коротко об интеграле можно сказать так:

интеграл — это площадь.

Способ вычисления площади, о котором пойдет речь в этой главе, уходит корнями в глубокую древность. Так, в III в. до н.э. великий Архимед вычислил площадь параболического сегмента с помощью изобретенного им «метода исчерпывания», который через 2000 лет преобразовался в метод интегрирования, применяемый к вычислению многих других величин.

Метод интегрирования для вычисления площади плоской фигуры требует, чтобы граница этой фигуры была задана аналитически, т.е. чтобы на плоскости была введена система координат, а граница была задана уравнением, связывающим координаты ее точек. Простейшими фигурами такого рода являются криволинейные трапеции.

Пусть на координатной плоскости дан график неотрицательной функции f, заданной на отрезке $[a, b]$. Криволинейной трапецией называется фигура, ограниченная графиком функции f, прямыми $x = a$ и $x = b$ и осью абсцисс (рис. 330).

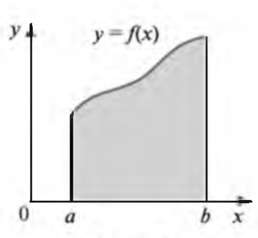


Рис. 330

Можно образовать криволинейные трапеции с помощью различных известных вам функций. Некоторые примеры представлены на рис. 331, *а*, *б*, *в*.

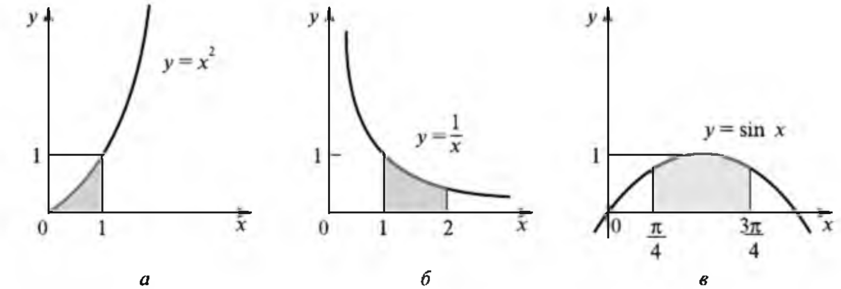


Рис. 331

Криволинейную трапецию, построенную с помощью графика неотрицательной функции *f*, будем сокращенно называть подграфиком функции *f*. Определим интеграл.

Пусть дана неотрицательная функция *f*, определенная на конечном отрезке [*a*, *b*]. Интегралом от функции *f* называется площадь ее подграфика.

Итак, *интеграл — это площадь.* Если мы научимся вычислять площади, то сумеем вычислить и интегралы, а значит, и многие физические величины.

Прямое вычисление площадей различных фигур (интегралов от некоторых функций) провел еще Архимед. Однако лишь в XVII в. Ньютону и Лейбницу удалось открыть общий способ вычисления интегралов.

Интегральные суммы

Метод исчерпывания Архимеда хотя и не дал общего способа вычисления площади, однако сыграл очень большую роль в математике, так как с его помощью удалось объединить самые разные задачи — вычисление площади, объема, массы, работы, давления, электрического заряда, светового потока и многие, многие другие.

Проиллюстрируем этот метод на простом примере. Предположим, что нам надо вычислить объем лимона. Он имеет неправильную форму, и применить какую-либо известную формулу нельзя. С помощью взвешивания найти объем также трудно, так как плотность лимона в разных частях его разная. Поступим следующим образом. Нарежем лимон на тонкие дольки. Каждую дольку приближенно можно считать цилиндриком, в основании его — круг, радиус которого можно измерить. Объем такого цилиндра вычислить легко по готовой формуле. Сложив объемы маленьких цилиндров, мы получим приближенное значение объема всего лимона. Приближение будет тем точнее, чем на более тонкие части мы сможем нарезать лимон.

Объясним более точно эту процедуру для вычисления площади подграфика. Рассмотрим подграфик функции f, заданной на отрезке $[a, b]$. Разобьем этот отрезок на несколько частей. Площадь всего подграфика разобьется на сумму площадей более мелких криволинейных трапеций. Каждую такую трапецию можно приближенно считать прямоугольником. Сумма площадей этих прямоугольников дает приближенное представление обо всей площади подграфика. Чем мельче мы разобьем отрезок $[a, b]$, тем точнее вычислим площадь.

Запишем проведенное рассуждение в виде формул.

Разделим отрезок $[a, b]$ на n частей $x_0 = a, x_1, ..., x_n = b$. Длину i-го отрезка обозначим через $\Delta x_i = x_i - x_{i-1}$. Составим сумму: $S_n = f(x_1)\,\Delta x_1 + ... + f(x_n)\,\Delta x_n$. Геометрически эта сумма представляет собой площадь ступенчатой заштрихованной фигуры (рис. 332).

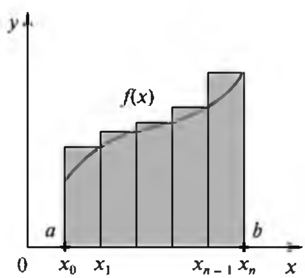


Рис. 332

Суммы вида $S_n = f(x_1)\,\Delta x_1 + ... + f(x_n)\,\Delta x_n$ называются *интегральными суммами для функции f*.

Интегральные суммы дают приближенное значение площади. Точное значение получается с помощью предельного перехода. Представим себе, что мы измельчаем разбиение отрезка $[a, b]$ так, что длины всех маленьких отрезков стремятся к нулю (т.е. $\Delta x_i \to 0$). Тогда площадь ступенчатой фигуры будет приближаться к площади подграфика S. Можно сказать, что площадь подграфика равна пределу интегральных сумм, т.е.

$$S = \lim S_n.$$

Интеграл можно определить так:

интеграл — это предел интегральных сумм.

С помощью интегральных сумм можно приближенно вычислять самые различные величины.

Примеры

1. *Объем лимона*. Обозначим Δh_i, толщину i-й дольки (не обязательно резать лимон на дольки одинаковой толщины), r_i ($i = 1, \dots, n$ — число долек) — ее радиус. Объем лимона приближенно представится интегральной суммой: $S_n = \pi r_1^2 \Delta h_1 + \dots + \pi r_n^2 \Delta h_n$.
2. *Работа*. Предположим, что на точку, движущуюся по оси x, действует некоторая сила F, направленная по той же оси. Если сила F постоянна, то работа равна Fs, где s — путь, пройденный точкой. Предположим, что F изменяется от точки к точке и ее значение в каждой точке некоторого промежутка $[a, b]$ равно $F(x)$. Как найти работу A по перемещению точки из a в b?
   Разобьем отрезок $[a, b]$ на n отрезков. Будем приближенно считать, что на каждом малом отрезке сила постоянна. Если i-й отрезок образован точками $[x_{i-1}, x_i]$, то в качестве силы можно взять значение функции F в одной из точек этого отрезка, например в x_i. Работа на i-м отрезке пути приближенно представится как произведение $F(x_i)\,\Delta x_i$, а на всем отрезке — интегральной суммой:
   $$A_n = F(x_1)\,\Delta x_1 + \dots + F(x_n)\,\Delta x_n.$$
   Точное значение работы A получим, переходя к пределу $A = \lim A_n$.

Вычислять интегралы через пределы соответствующих интегральных сумм очень трудно. Архимед сумел найти некоторые площади, объемы фактически через пределы интегральных сумм для квадратичных функций. Однако этот результат стоял особняком в математике до конца XVII в., когда было выяснено, что нахождение площади является обратной задачей к задаче нахождения скорости.

Скорость роста площади

Рассмотрим неотрицательную функцию f, заданную на конечном отрезке $[a, b]$. Представим себе «переменную» криволинейную трапецию, полученную следующим образом: закрепим левую стенку $x = a$, а правую начнем перемещать вдоль оси абсцисс. Такая трапеция изображена на рис. 333. Ее можно считать подграфиком функции f, область определения которой ограничена отрезком $[a, x]$.

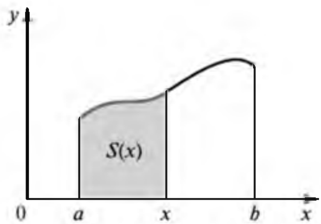


Рис. 333

Обозначая площадь переменной трапеции, т.е. площадь подграфика функции f, заданной на отрезке $[a, x]$, через $S(x)$, получаем новую функцию S (переменная площадь): $S = S(x)$. Перечислим свойства функции S. Она определена для всех $x \in [a, b]$, ее значение при $x = a$ равно нулю (трапеция вырождается в отрезок, и ее площадь равна нулю), эта функция возрастает, и при $x = b$ ее значение равно площади всего подграфика, т.е. интегралу от функции f.

Найдем скорость роста функции S и результат запишем в виде теоремы.

Теорема о скорости роста площади. Пусть f — неотрицательная функция, S — переменная площадь ее подграфика, тогда производная функции S равна функции f, т.е. $S'(x) = f(x)$.

Теорема утверждает, что производная переменной площади подграфика функции f равна самой функции f Для доказательства теоремы поступим так, как всегда поступают при вычислении производной.

Зафиксируем значение аргумента x и придадим аргументу приращение Δx. Вычислим приращение функции: $\Delta S = S(x + \Delta x) - S(x)$. Из рисунка 334 следует, что приращение площади есть площадь подграфика функции f, определенной на отрезке $[x, x + \Delta x)$.

Если Δx достаточно мало, то площадь криволинейной трапеции мало отличается от площади прямоугольника со сторонами $f(x)$ и Δx, т.е. можно записать приближенное равенство $\Delta S \approx f(x) \cdot \Delta x$. Следовательно, $\dfrac{\Delta S}{\Delta x} \approx f(x)$, т.е. производная функции S равна функции f, что и утверждалось в теореме.

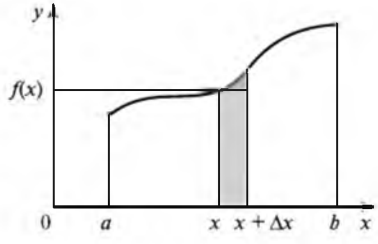


Рис. 334

В доказательстве теоремы мы использовали такую идею: если отрезок $[x, x + \Delta x]$ достаточно мал, то площадь подграфика функции f на этом отрезке мало отличается от произведения $f(x)\ \Delta x$. Но если в точке x функция f имеет разрыв, то это неверно, как хорошо видно из рис. 335. Поэтому в формулировку теоремы о скорости роста площади надо добавить требование непрерывности функции f.

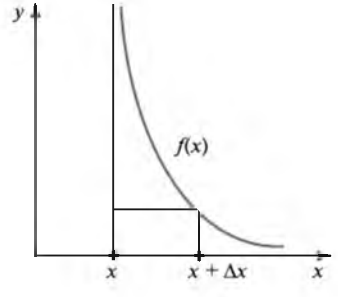


Рис. 335

Итак, для функции f мы получили новую функцию S — переменную площадь подграфика. Связь между функциями f и S такова:

S — интеграл от функции f,
f — производная функции S.

Из этого сравнения видно, что нахождение интеграла (интегрирование) и нахождение производной (дифференцирование) являются взаимно обратными операциями. Если известна функция f, то нахождение функции S (площади подграфика функции f) есть задача интегрирования функции f Если же задана функция S, то нахождение функции f (скорости роста площади) есть задача дифференцирования функции S.

Обозначение интеграла

Интеграл от функции $y = f(x)$ на отрезке $[a, b]$ обозначается так:

$$\int_a^b f(x)\mathrm{d}x .$$

Эта традиция имеет исторические корни. Интегральные суммы, с помощью которых приближенно вычисляется интеграл: составляются из слагаемых вида $f(x)\ \Delta x$. Приближенное равенство $\Delta S \approx f(x)\ \Delta x$ может быть заменено точным равенством дифференциалов $\mathrm{d}S = f(x)\ \mathrm{d}x$. Интеграл можно представить как сумму «бесконечного числа дифференциалов». Знак интеграла и есть стилизованная форма буквы S — первой буквы слова *summa* (лат.). Запись

$$S = \int_a^b f(x)\mathrm{d}x .$$

напоминает, что площадь S можно получить суммированием слагаемых вида $f(x)\ \mathrm{d}x$. Около знака интеграла ставят пределы интегрирования — концы отрезка $[a, b]$, на котором задана функция f.

Переменная площадь $S(x)$ (площадь подграфика функции f на отрезке $[a, x]$) запишется в виде интеграла с переменным верхним пределом:

$$S(x) = \int_a^x f(x)\mathrm{d}x .$$

Связь между функциями f и S, установленную в теореме о скорости роста площади, можно записать так:

$$\left(\int_a^x f(x)\mathrm{d}x\right)' = f(x).$$

Отсюда видно, что операции интегрирования и дифференцирования являются взаимно обратными.

Определение интеграла и полученные для него результаты нетрудно распространить на произвольную функцию f, отказавшись от требования ее неотрицательности. Рассмотрим произвольную функцию, заданную на отрезке $[a, b]$, и ее подграфик. т.е. часть плоскости, ограниченную графиком f, прямыми $x = a$, $x = b$ и осью абсцисс (рис. 336).

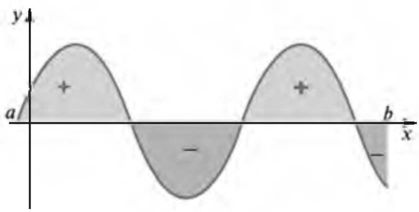


Рис. 336

Интегралом от функции называется сумма площадей частей ее подграфика, взятых со знаком «+» для частей, расположенных выше оси абсцисс, и со знаком «–» — ниже ее.

Контрольные вопросы и задания

1. Обратите внимание на следующие встретившиеся в тексте слова и обозначения: подграфик, интеграл, интегральная сумма, интеграл с переменным верхним пределом. Приведите примеры их использования.
2. Какие физические величины можно выразить как площади?
3. Что такое подграфик функции?
4. Что называется интегралом?
5. Как строится интегральная сумма?
6. Что такое интеграл с переменным верхним пределом?
7. В чем состоит связь между функцией и переменной площадью ее подграфика?

8. Как определить интеграл от функции, принимающей значения разных знаков?
9. Какое условие нужно наложить на функцию для справедливости теоремы о скорости роста площади?
10. Чему равен дифференциал переменной площади подграфика функции f?

§ 33. Вычисление интеграла

Первообразная функция

В предыдущем параграфе мы установили, что интегрирование является операцией, в определенном смысле обратной дифференцированию. Вычисление интегралов сводится к нахождению функции, производная которой равна заданной функции. Эту операцию мы рассмотрим отдельно.

Первообразной для функции f называется такая функция F, производная которой равна данной функции.

Иными словами, равенство

$$F' = f$$

можно прочесть двояко: f — производная функции F, или F — первообразная для функции f Для обозначения первообразной используют знак «неопределенного интеграла», т.е. интеграла без указания пределов интегрирования:

$$F = F = \int f(x)\mathrm{d}x\,.$$

Свойства первообразной функции.

1. Если F — первообразная для функции f, то $F + C$, где C — константа, также первообразная для той же функции f.

Действительно, $(F + C)' = F' + C' = f + 0 = f$.

2. Обратно: если F_1 и F_2 — две первообразные для одной и той же функции f, то они отличаются на постоянное слагаемое.

Действительно, из $F_1' = f$ и $F_2' = f$ следует $(F_1 - F_2)' = F_1' - F_2' = f - f = 0$. Функция, производная которой тождественно равна нулю, является постоянной. Это свойство производной очевидно из ее механического смысла: если скорость точки равна нулю, то она стоит на месте и ее координата постоянна. Итак, $F_1 - F_2 = C$. Все первообразные для функции f получаются из одной из них прибавлением к ней произвольной постоянной. Надо помнить, что знак $\int$ является «неопределенным» в том смысле, что он обозначает какую-либо первообразную.

3. $\int\left(f(x)+g(x)\right)\mathrm{d}x=\int f(x)\mathrm{d}x+\int g(x)\mathrm{d}x.$

Действительно, пусть F и G — первообразные для функций f и g соответственно. Тогда $F + G$ является первообразной для функции $f + g$: $(F + G)' = F' + G' = f + g$.

4. $\int cf(x)\mathrm{d}x=c\int f(x)\mathrm{d}x.$

Доказывается аналогично.

Таблицу первообразных получают с помощью таблицы производных. Проверить табл. 6 можно, делая обратную операцию, т.е. вычисляя производные.

Таблица 6

Первообразные функции

$f(x)$	$\int f(x)\mathrm{d}x$
1	x
x	$\frac{x^2}{2}$
xk	$\frac{x^{k+1}}{k+1}$ $(k \neq 1)$
$\frac{1}{x}$	$\lambda\nu\, x\ (x > 0)$
$\sin x$	$-\cos x$
$\cos x$	$\sin x$
ex	ex
$\frac{1}{1+x^2}$	$\mathrm{arctg}\, x$

Линейная замена переменной

Теорема. Пусть F — первообразная для функции f. Тогда

$$\int f(kx+b)\,\mathrm{d}x = \frac{1}{k}F(kx+b)\,.$$

Действительно, вычислим производную от $F(kx + b)$: $(F(kx + b))' = kF'(kx + b) = kf(kx + b)$. Отсюда $\frac{1}{k}F(kx+b)$ является первообразной для функции $f(kx + b)$.

Следовательно, таблица первообразных может быть дополнена (табл. 7).

Таблица 7

Первообразные функции (дополнение)

$f(x)$	$\int f(x)\,\mathrm{d}x$
$(kx + b)^n$	$\frac{1}{k}\frac{(kx+b)^{n+1}}{n+1}\ (n \neq -1)$
$\frac{1}{x-\alpha}$	$\ln(x-\alpha)\ (x > \alpha)$
$a^x = e^{(\ln a)x}$	$\frac{1}{\ln a}a^x$
$\sin(\omega x + \alpha)$	$-\frac{1}{\omega}\cos(\omega x + \alpha)$
$\cos(\omega x + \alpha)$	$\frac{1}{\omega}\sin(\omega x + \alpha)$

Операция дифференцирования совершается формально — нужно запомнить несколько правил и их будет достаточно для нахождения производных. Не так обстоит дело с интегрированием: например, нет формулы для интегрирования произведения и частного функций. Поэтому составлены обширные таблицы интегралов (первообразных) и появляется новая задача — научиться преобразовывать вычисляемые интегралы к табличным.

Пример

Вычислить $\int \sin^2 x\,\mathrm{d}x$.

В таблице интегралов нет интеграла от $\sin^2 x$. Однако можно воспользоваться формулой $\sin^2 x = \frac{1}{2}(1-\cos 2x)$. Для $\cos 2x$ интеграл известен, поэтому

$$\int \sin^2 x\mathrm{d}x = \int \frac{1}{2}(1-\cos 2x)\mathrm{d}x = \frac{1}{2}\left(\int \mathrm{d}x - \int \cos 2x\mathrm{d}x\right) =$$
$$= \frac{1}{2}\left(x - \frac{1}{2}\sin 2x\right) = \frac{x}{2} - \frac{\sin 2x}{4}.$$

Теорема Ньютона — Лейбница

Знаменитая теорема, названная именами основоположников математического анализа, гласит:

интеграл равен приращению первообразной.

Запишем формулировку более подробно.

Теорема Ньютона — Лейбница. Пусть f — данная функция, F — ее первообразная. Тогда

$$\int_a^b f(x)\mathrm{d}x = F(b) - F(a)$$

Доказательство. Одна первообразная для функции f известна — это $S(x)$ — переменная площадь подграфика функции f. По определению интеграла $\int_a^b f(x)\mathrm{d}x = S(b) - S(a) = S(b)$, так как $S(a) = 0$. Пусть F — произвольная первообразная для функции f. Тогда она отличается от S на константу, но приращение функций F и S будет одним и тем же:

$$S(x) = F(x) + C,$$

$$\int_a^b f(x)\,\mathrm{d}x = S(b) - S(a) = (F(b)+C) - (F(a)+C) = F(b) - F(a),$$

что и требовалось доказать. Разность $F(b) - F(a)$ сокращенно записывается в виде $F(x)\big|_a^b$.

Теорема Ньютона — Лейбница сводит вычисления интегралов к вычислению первообразных.

Примеры

1. Вычислить $\int\limits_0^1 x^2 dx$ (рис. 337).

$$\int\limits_0^1 x^2 dx = \frac{x^3}{3}\Big|_0^1 = \frac{1}{3} - 0 = \frac{1}{3}.$$

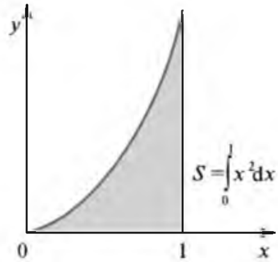


Рис. 337

2. Вычислить $\int\limits_0^{\pi/2} \sin 2x\, dx$.

$$\int\limits_0^{\pi/2} \sin 2x\, dx = -\frac{1}{2}\cos 2x\Big|_0^{\pi/2} = -\frac{1}{2}\cos\pi + \frac{1}{2}\cos 0 = 1.$$

Свойства интеграла

Формула Ньютона — Лейбница сводит свойства интеграла к свойствам первообразной, которые, в свою очередь, опираются на свойства производной.

1. $\int\limits_a^b (f+g)\, dx = \int\limits_a^b f\, dx + \int\limits_a^b g\, dx$.

2. $\int\limits_a^b cf\, dx = c\int\limits_a^b f\, dx$.

Доказательство (первого свойства). Пусть F и G — первообразные для функций f и g соответственно. Тогда функция $F + G$ — одна из первообразных функций $f + g$. По теореме Ньютона — Лейбница,

$$\int\limits_a^b (f+g)\, dx = (F+G)\Big|_a^b = F(b) + G(b) - F(a) - G(a) =$$

$$= F(b) - F(a) + G(b) - G(a) = \int\limits_a^b f\, dx + \int\limits_a^b g\, dx.$$

Второе свойство доказывается аналогично. Отметим следствия свойств 1 и 2:

$$\int\limits_a^b (f-g)\, dx = \int\limits_a^b f\, dx - \int\limits_a^b g\, dx,$$

$$\int\limits_a^b (\alpha f + \beta g)\, dx = \alpha\int\limits_a^b f\, dx + \beta\int\limits_a^b g\, dx.$$

Некоторые свойства интеграла являются следствием свойств площади, лежащей в определении интеграла.

3. $\int\limits_a^b f\mathrm{d}x = \int\limits_a^c f\mathrm{d}x + \int\limits_c^b f\mathrm{d}x$.

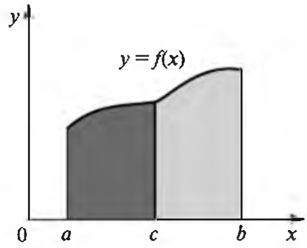


Рис. 338

Доказательство. Площадь всей криволинейной трапеции с основанием $[a, b]$ (рис. 338) есть сумма площадей трапеций с основаниями $[a, c]$ и $[c, b]$.

Это же свойство можно получить и вычислением. Пусть F — первообразная для функции f_c. Тогда

$$\int\limits_a^c f\,\mathrm{d}x = F(c) - F(a), \int\limits_c^b f\,\mathrm{d}x = F(b) - F(c).$$

Складывая почленно последние равенства, получаем

$$F(c) - F(a) + F(b) - F(c) = F(b) - F(a) = \int\limits_a^b f\,\mathrm{d}x.$$

Доказанное свойство интеграла называют его аддитивностью (от лат. *addo* — складываю).

Полезно отметить, что $\int\limits_a^a f(x)\mathrm{d}x = 0$, так как $F(a) - F(a) = 0$.

4. Если $f(x) \ge g(x)$, то $\int\limits_a^b f(x)\mathrm{d}x \ge \int\limits_a^b g(x)\mathrm{d}x$.

Действительно, функция $h(x) = f(x) - g(x)$, по условию, не отрицательна. Следовательно, не отрицателен от нее и интеграл, являющийся, по определению, площадью подграфика: $\int\limits_a^b (f(x) - g(x))\,\mathrm{d}x \ge 0$. Раскрывая левую часть этого неравенства по свойствам 1 и 2, получаем $\int\limits_a^b f(x)\mathrm{d}x - \int\limits_a^b g(x)\,\mathrm{d}x \ge 0$ т.е. $\int\limits_a^b f(x)\mathrm{d}x \ge \int\limits_a^b g(x)\,\mathrm{d}x$, что и требовалось доказать.

Примеры

1. Вычислить $\int\limits_0^{\pi/2} (x - \sin x)\mathrm{d}x$.

$$\int_0^{\pi/2}(x-\sin x)\,dx=\int_0^{\pi/2}x\,dx-\int_0^{\pi/2}\sin x\,dx=\frac{x^2}{2}\Big|_0^{\pi/2}+\cos x\Big|_0^{\pi/2}=\frac{\pi^2}{8}-1\approx 0{,}23$$

2. $$\int_1^2\frac{x^2+x+1}{x^2}dx=\int_1^2\left(1+\frac{1}{x}+\frac{1}{x^2}\right)dx=x\Big|_1^2+\ln x\Big|_1^2-\frac{1}{x}\Big|_1^2=$$

$$=2-1+\ln 2-0-\frac{1}{2}+1=\frac{3}{2}+\ln 2\approx 2{,}19.$$

Контрольные вопросы и задания

1. Что такое первообразная?
2. Перечислите свойства первообразной.
3. Как меняется первообразная при линейной замене аргумента подынтегральной функции?
4. Сформулируйте теорему Ньютона — Лейбница.
5. Перечислите свойства интеграла.
6. В чем состоит аддитивность интеграла?
7. Как связаны между собой две первообразные для одной и той же функции?
8. Как меняется при интегрировании амплитуда гармонического колебания?
9. Как меняется степень при интегрировании степенной функции?
10. Верно ли, что интеграл от любой степенной функции будет снова степенной функцией?

§ 34. Приложения интеграла

Площадь

Пусть надо вычислить площадь какой-либо плоской фигуры Ф. Введем на плоскости декартову систему координат. Тогда отдельные куски границы фигуры Ф можно задать в виде графиков некоторых функций. Сравнительно простой случай изображен на рис. 339. Площадь фигуры Ф можно получить сложением и вычитанием площадей подграфиков функций, задающих границу.

Рис. 339

Так, в типичном случае площадь фигуры Φ (рис. 340) получается как разность подграфиков функций f и g, т.е. выражается через интегралы:

$$S = \int_a^b f(x)\,\mathrm{d}x - \int_a^b g(x)\,\mathrm{d}x\ .$$

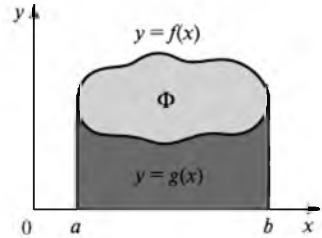


Рис. 340

Примеры

1. Найти площадь одной арки синусоиды (рис. 341).

$$S = \int_0^{\pi} \sin x\,\mathrm{d}x = -\cos x\Big|_0^{\pi} = 1 + 1 = 2$$

Найти площадь между дугами парабол $y = x^2$ и $y = \sqrt{x}$ (рис. 342).

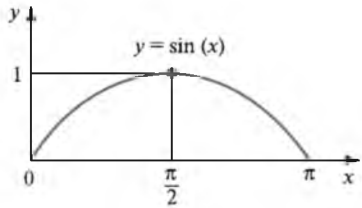


Рис. 341

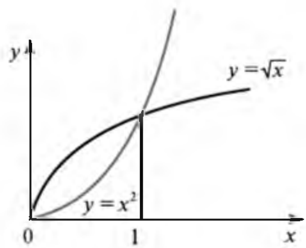


Рис. 342

Данная фигура ограничена графиками двух функций: $f(x) = \sqrt{x}$ и $g(x) = x^2$. Искомая площадь равна

$$S = \int_0^1 \sqrt{x}\,\mathrm{d}x - \int_0^1 x^2\mathrm{d}x = \frac{x^{3/2}}{3/2}\Bigg|_0^1 - \frac{x^3}{3}\Bigg|_0^1 = \frac{2}{3} - \frac{1}{3} = \frac{1}{3}.$$

То, что площадь вычисляется с помощью интеграла, — это не удивительно, так как понятие интеграла с самого начала у нас вводилось через понятие площади. Вычисление других физических величин с помощью интеграла потребует дополнительных рассуждений.

Схема применения интеграла

При ознакомлении с понятием интеграла мы выделили три его характеристики.

1. Интеграл от функции f есть площадь ее подграфика (с учетом знака).
2. Интеграл есть предел интегральных сумм.
3. Интеграл от функции f есть приращение ее первообразной.

Любая из этих характеристик интеграла может служить основой для его приложений. Наиболее стандартным путем выражения какой-либо физической величины в виде интеграла от другой является использование третьей характеристики интеграла как приращения первообразной. Однако и две первые характеристики очень важны в приложениях, так как позволяют получить геометрический смысл связи между физическими величинами и дать простой способ приближенного вычисления.

Вернемся еще раз к величинам, которые вычисляются с помощью интеграла (например, такие физические понятия, как перемещение, работа, масса, электрический заряд, давление, теплота). К ним можно добавить геометрические величины — длину, площадь, объем. Все эти величины обладают некоторыми общими свойствами.

1. Величины можно рассматривать как функции отрезка.

Перемещение вычисляется в зависимости от отрезка времени движения. Работа переменной силы при движении по прямой зависит от пройденного отрезка пути. Массу тонкого неоднородного стержня можно рассматривать как функцию от отрезков этого стержня. Электрический заряд, протекающий через поперечное сечение проводника, зависит от отрезка времени, за который производится измерение.

Функции отрезка часто можно представить как приращение обычной функции. Так, при движении точки по прямой, можно рассматривать эту прямую как числовую ось и ввести функцию времени — координату точки. Тогда перемещение на данном отрезке времени запишем как приращение координаты:

$$x = x(t),\ S = \Delta x = x(t_2) - x(t_1).$$

При вычислении массы стержня можно считать, что стержень расположен на числовой оси x и занимает положение от 0 до l (l — длина стержня). Теперь введем функцию $m = m(x)$ — массу стержня с концами в начальной точке и в точке x. Тогда масса отрезка стержня с концами в точках x_1 и x_2 равна приращению функции $m(x)$ на этом отрезке:

$$\Delta m = m(x_2) - m(x_1).$$

Для вычисления площади подграфика мы вводили переменную площадь $S = S(x)$ — площадь под графиком функции от начальной точки до точки x. Площадь криволинейной трапеции с основанием $[x_1, x_2]$ равна приращению функции S на этом отрезке.

Заметим, что так как мы вычисляли приращение функции, а оно не изменится, если функцию изменить на постоянное слагаемое, то при определении переменных массы, работы или площади не важно, какое значение мы выбираем для этой функции в начальной точке или от какой точки начинаем вычислять ее значения.

2. Известна скорость изменения величины.

Так, скоростью изменения перемещения будет скорость в обычном смысле этого слова. Скоростью изменения работы в зависимости от времени является мощность, а скоростью изменения работы в зависимости от перемещения является сила. Скорость изменения массы — это плотность. Само слово «плотность» имеет такой же универсальный характер, как слово «скорость», и им широко пользуются. Например, плотность тока, плотность заряда.

Если величину записать как приращение некоторой обычной функции, то ее скорость (или плотность) равна производной этой функции.

Примеры

1. Скорость механического движения

$$v = \frac{\mathrm{d}x}{\mathrm{d}t}.$$

2. Линейная плотность стержня

$$\rho = \frac{\mathrm{d}m}{\mathrm{d}x}.$$

3. Мощность

$$N = \frac{\mathrm{d}A}{\mathrm{d}t}.$$

4. Сила при перемещении по прямой

$$F = \frac{\mathrm{d}A}{\mathrm{d}x}.$$

Интеграл применяется тогда, когда известна скорость или плотность f искомой величины. Если искомая величина представлена в виде приращения некоторой функции F, то f является производной F, а F — первообразной для функции f, т.е. интегралом от функции f.

Запишем этот вывод с помощью формул. Независимый аргумент обозначим t, а искомую величину — F. Рассмотрим значение F на малом отрезке $[t, t + \mathrm{d}t]$. Пусть f — скорость изменения F. Запишем связь между F и f в дифференциальной форме:

$$\mathrm{d}F = f(t)\,\mathrm{d}t.$$

Следовательно,

$$F = \quad F = \int_a^b f(t)\mathrm{d}t \cdot$$

Итак, *схема применения интеграла* сводится к следующему.

1. Записываем главную часть изменения искомой величины с помощью дифференциалов: $\mathrm{d}F = f(t)\,\mathrm{d}t$.
2. Запишем F в виде интеграла:

$$F = \quad F = \int_a^b f(t)\mathrm{d}t \cdot$$

3. Находим первообразную для функции f и вычисляем F как разность значений первообразной на концах отрезка.

1. Отметим вольность в выборе обозначений, которая часто допускается в приложениях. Мы ищем некоторую величину, например массу стержня. Обозначаем ее какой-то буквой, скажем, m. В данном случае m — масса определенного стержня, т.е. просто число. Однако для нахождения m надо рассматривать массу произвольного отрезка стержня. Ее снова обозначаем m, хотя сейчас понимаем m как функцию отрезка. Затем мы вводим переменную массу — массу стержня от начальной точки до точки x — и эту функцию от x тоже удобно обозначать той же буквой m, чтобы записать соотношение для плотности: $\rho = \dfrac{\mathrm{d}m}{\mathrm{d}x}$, или $\mathrm{d}m = \rho\mathrm{d}x$. Мы употребляем везде одну и ту же букву (хотя речь идет о разных понятиях, для обозначения используется одна и та же буква).

2. Рассмотренные нами величины были функциями отрезка. Часто встречаются аналогичные величины, но зависящие от других областей. Так, масса произвольного тела зависит не от отрезка, как это мы идеализировали для тонкого стержня, а от области пространства. Существуют более сложные интегралы, которые позволяют вычислить величины, зависящие от частей кривой линии (криволинейные интегралы), от частей поверхности (поверхностные интегралы), от частей объема (объемные, или тройные, интегралы). В некоторых простых случаях искомые величины удается рассмотреть как функции отрезка и свести их нахождение к вычислению обычных интегралов. С такими примерами мы ознакомимся при вычислении объемов и при решении прикладных задач.

Работа

Пусть точка движется по оси x, в каждой точке которой приложена некоторая сила $F = F(x)$. Вычислим работу, которую надо совершить при перемещении из точки a в точку b. На малом отрезке пути от точки x до точки $x + \mathrm{d}x$ можно считать силу постоянной и равной $F(x)$. Тогда дифференциал работы $\mathrm{d}A = F(x)\mathrm{d}x$. Отсюда получаем, что вся работа на отрезке $[a, b]$ равна

$$A = \int_a^b F(x)\mathrm{d}x \cdot$$

Эта формула позволяет вычислить работу при прямолинейном движении. Перемещение по криволинейным траекториям требует другого, так называемого криволинейного интеграла, однако запись работы в виде интеграла осталась бы прежней. В этом случае сила была бы векторной величиной $\mathbf{F}$, перемещение совершалось бы не обязательно по прямой, и на малом отрезке его можно было бы записать $\mathrm{d}\mathbf{r}$. Главная часть работы есть скалярное произведение $\mathrm{d}A = \mathbf{F}\mathrm{d}\mathbf{r}$, а вся работа — криволинейный интеграл $\int\limits_{\Gamma} \mathbf{F}\mathrm{d}\mathbf{r}$, где Г — кривая, по которой происходит перемещение точки (рис. 343).

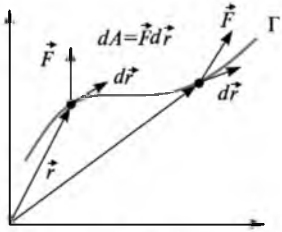


Рис. 343

Пример

Предположим, что в точке O помещен единичный электрический заряд. Он создает электрическое поле. На другой единичный заряд, помещенный в точку x, действует сила F, обратно пропорциональная квадрату расстояния, т.е. $F(x) = \dfrac{k}{x^2}$. Найти работу электрического поля по перемещению единичного заряда из точки x_1 в точку x_2. Применяя формулу для работы, получаем

$$A = \int_{x_1}^{x_2} \frac{k}{x^2}\mathrm{d}x = k\int_{x_1}^{x_2} \frac{\mathrm{d}x}{x^2}.$$

Для функции $F(x) = \dfrac{k}{x^2}$ первообразную $u(x)$ можно найти по таблице: $u(x) = -\dfrac{k}{x}$. Тогда $A = u(x)\Big|_{x_1}^{x_2} = \left(-\dfrac{k}{x}\right)\Big|_{x_1}^{x_2} = \dfrac{k}{x_1} - \dfrac{k}{x_2}$.

Функция $u(x) = -\dfrac{k}{x}$ называется *потенциалом электрического поля*. Работа A равна приращению функции u, т.е. разности потенциалов на концах отрезка.

Поскольку геометрический смысл интеграла — это площадь, работу можно представить как площадь подграфика для функции $F(x)$. Изобразим этот график для разобранного примера (рис. 344).

Перемещение

Предположим, что точка движется по прямой (по оси x) с известной скоростью. Положение точки на оси будем считать функцией времени: $x = x(t)$. Как найти перемещение точки за промежуток времени $[t_1, t_2]$?

Если скорость точки постоянна и равна v, то перемещение $S = v(t_2 - t_1)$. Пусть теперь скорость меняется по закону $v = v(t)$. За промежуток времени $[t, t + \Delta t)$ главную часть перемещения Δs мы получим, если будем считать, что на этом промежутке скорость постоянна и равна $v(t)$. Тогда $\mathrm{d}S = v(t)\mathrm{d}t$; следовательно, $S = \int\limits_{t_1}^{t_2} v(t)\mathrm{d}t$.

Если изобразить график скорости, то *перемещение* будет задаваться площадью подграфика (рис. 345).

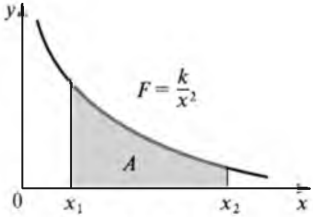


Рис. 344

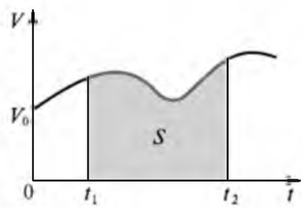


Рис. 345

Масса

Масса произвольного тела является величиной, для вычисления которой нужно тоже понятие интеграла, только более сложное. Мы сможем написать формулу для вычисления массы тонкого стержня, т.е. такого тела, в котором плотность изменяется вдоль одного направления и его можно представить как отрезок тонкой проволоки с изменяющейся плотностью.

Если стержень однороден, то его масса m пропорциональна длине l, т.е. $m = \rho l$, где ρ — коэффициент пропорциональности, называемый *линейной плотностью.* Вычислим массу неоднородного стержня, если известно, как меняется плотность ρ. Представим себе, что стержень расположен вдоль оси x так, что занимает положение $[0, l]$. Если линейную плотность ρ можно считать функцией от x: $\rho = \rho(x)$, заданной на этом отрезке, то, считая на отрезке $[x, x + \mathrm{d}x]$ плотность постоянной, получаем $\mathrm{d}m = \rho(x)\,\mathrm{d}x$, следовательно,

$$m = \int_0^l \rho(x)\mathrm{d}x .$$

Таким образом, масса стержня является интегралом от его линейной плотности.

Электрический заряд

Представим себе переменный ток, текущий по проводнику. Как вычислить заряд q, переносимый за интервал времени $[a, b]$ через сечение проводника? Если сила тока I не изменяется со временем, то изменение заряда q равно $I(b - a)$. Пусть задан закон изменения $I = I(t)$ в зависимости от времени. На малом интервале времени $[t, t + \mathrm{d}t]$ можно считать силу тока постоянной и равной $I(t)$, тогда $\mathrm{d}q = I(t)\,\mathrm{d}t$ и, следовательно,

$$q = \int_a^b I(t)\mathrm{d}i .$$

Сравним эти примеры с физическими примерами, обсуждавшимися в связи с производной. Фактически рассматриваются одни и те же соотношения вида $\mathrm{d}F = f(x)\,\mathrm{d}x$, но в первом случае дана величина F и нужно найти f (f выступает как производная F), а во втором случае известна величина f и надо найти F (F является интегралом от f).

Составим из наших примеров табл. 8

Прикладные задачи

Пирамида Хеопса представляет собой правильную четырехугольную пирамиду высотой 147 м, в основании которой квадрат со стороной 232 м. Она построена из камня, плотность которого 2,5 г/см^3. Найти работу против силы тяжести, затраченную при постройке.

Таблица 8

Приложения интеграла

Величина	Соотношениев дифференциалах	Производная	Интеграл
A — работа N — мощность	$\mathrm{d}A = N(t)\,\mathrm{d}t$	$N(t) = \dfrac{\mathrm{d}A}{\mathrm{d}t}$	$A = \int_{t_1}^{t_2} N(t)\mathrm{d}t$
F — сила	$\mathrm{d}A = F(x)\,\mathrm{d}x$	$F(x) = \dfrac{\mathrm{d}A}{\mathrm{d}x}$	$A = \int_{x_1}^{x_2} F(x)\mathrm{d}x$
m — масса тонкого стержня ρ — линейная плотность	$\mathrm{d}m = \rho(x)\,\mathrm{d}x$	$\rho(x) = \dfrac{\mathrm{d}m}{\mathrm{d}x}$	$m = \int_{x_1}^{x_2} \rho(t)\mathrm{d}t$
q — электрический заряд I — сила тока	$\mathrm{d}q = I(t)\,\mathrm{d}t$	$I(t) = \dfrac{\mathrm{d}q}{\mathrm{d}t}$	$q = \int_{t_1}^{t_2} I(t)\mathrm{d}t$
s — перемещение v — скорость	$\mathrm{d}s = v(t)\,\mathrm{d}t$	$v(t) = \dfrac{\mathrm{d}s}{\mathrm{d}t}$	$s = \int_{t_1}^{t_2} v(t)\mathrm{d}t$

Проведем вертикально вверх ось x с началом у основания пирамиды. По этой оси будем измерять высоту подъема камней.

Сначала решим задачу в общем виде. Обозначим: H — высота пирамиды; a — сторона основания; ρ — плотность камня, $A(x)$ — работа, выполненная при возведении пирамиды от основания до высоты x. Найдем сначала сторону у квадрата, получающегося в горизонтальном сечении пирамиды по высоте x. Из подобия треугольников получаем $\dfrac{H-x}{H} = \dfrac{y}{a}$, откуда $y = \dfrac{a}{H}(H-x)$ (рис. 346).

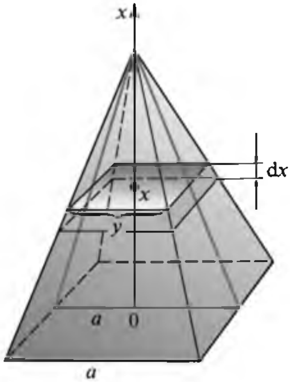


Рис. 346

Рассмотрим тонкий слой пирамиды, расположенный на расстоянии x от основания. Пусть толщина слоя равна $\mathrm{d}x$. Слой можно приблизительно считать параллелепипедом. Его масса $\mathrm{d}m = \rho y^2\,\mathrm{d}x = \rho\dfrac{a^2}{H^2}(H-x)^2\mathrm{d}x$. При подъеме этого слоя на высоту x была выполнена работа $\mathrm{d}A = \mathrm{d}mgx$, где g — ускорение падения, т.е. $\mathrm{d}A = \rho g\dfrac{a^2}{H^2}x(H-x)^2\,\mathrm{d}x$. Отсюда

$$A = A(H) = \int_0^H \rho g \frac{a^2}{H^2} x(H-x)^2 \mathrm{d}x = \frac{\rho g a^2}{H^2} \int_0^H (xH^2 - 2Hx^2 + x^3) =$$

$$= \frac{\rho g a^2}{H^2}\left(\frac{H^2 x^2}{2} - 2H\frac{x^3}{3} + \frac{x^4}{4}\right)\Bigg|_0^H = \frac{\rho g a^2 H^2}{12} = 2{,}37 \cdot 10^{12} \text{ Дж.}$$

Контрольные вопросы и задания

1. Как вычислить площадь плоской фигуры с помощью интеграла?
2. Назовите примеры физических величин, которые можно рассматривать как функции отрезка.
3. В чем состоит схема применения интеграла?
4. Что является плотностью работы как функции отрезка пути?
5. Что является плотностью перемещения как функции отрезка времени?
6. Что является плотностью массы тонкого стержня как функции отрезка этого стержня?
7. Что является плотностью электрического заряда как функции отрезка времени?
8. Что является плотностью количества теплоты как функции отрезка времени?
9. Как в форме интеграла записать дифференциальную связь $\mathrm{d}F = f(t)\,\mathrm{d}t$ между величинами F, f и t?
10. Как вычислить некоторую величину, если задан график ее плотности?

§ 35. Площади плоских фигур

Аксиомы площади

Площадь — это положительная скалярная величина. Если зафиксировать единицу площади, то площадь каждой измеряемой фигуры будет неотрицательным числом. Отметим это в качестве первой аксиомы площади.

Аксиома S_1 (неотрицательность площади)

Площадь каждой измеряемой плоской фигуры есть неотрицательное число.

Обозначим фигуру F, ее площадь $S(F)$. Тогда аксиому S_1 можно записать в виде неравенства $S(F) \geq 0$.

На самом деле аксиома S_1 подразумевает несколько больше, чем просто утверждение о неотрицательности площади. В ней неявно говорится о том, что всякой измеряемой фигуре можно сопоставить некоторое неотрицательное число, т.е. о задании отображения S множества измеряемых фигур во множество неотрицательных чисел. Следующие аксиомы фиксируют основные свойства этого отображения.

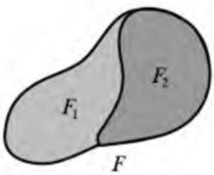


Рис. 347

Аксиома S_2 (аддитивность площади)

Пусть фигура F разбита на две части F_1 и F_2, не имеющие общих точек (рис. 347). Тогда площадь фигуры F равна сумме площадей фигур F_1 и F_2.

Используя теоретико-множественную символику, можно записать вторую аксиому площади так: если $F = F_1 \cup F_2$ (причем $F_1 \cap F_2 = \varnothing$), то $S(F_1) + S(F_2) = S(F)$.

Аксиома S_3 (нормированность площади)

Площадь прямоугольника равна произведению длин сторон.

Если F — прямоугольник со сторонами a и b, то $S(F) = ab$.

Аксиома S_4 (инвариантность площади)

Равные фигуры имеют равные площади.

Если фигуры F_1 и F_2 равны, т.е. совмещаются при перемещении, то их площади равны.

Остальные свойства площади можно вывести из аксиом $S_1 - S_4$. Для примера докажем свойство монотонности площади.

Теорема. Площадь части фигуры не больше площади всей фигуры.

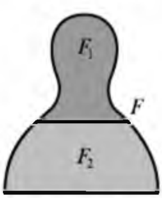


Рис. 348

Доказательство. Пусть фигура F_1 является частью фигуры F (рис. 348). Обозначим F_2 ту часть фигуры F, которая состоит из точек, не входящих в F_1. Тогда фигура F разбивается на две части F_1 и F_2, не имеющие общих точек. По аксиоме S_2 имеем $S(F) = S(F_1) + S(F_2)$. Слагаемое $S(F_2)$ не отрицательно по аксиоме S_1, следовательно, $S(F) \geq S(F_1)$, что и составляет утверждение этой теоремы.

Обратите внимание на неточность рис. 347, иллюстрирующего аксиому S_2. Фигура F разбита на две части: F_1 и F_2. Условие аксиомы требует, чтобы F_1 и F_2 не имели общих точек. Но как быть с границей между F_1 и F_2? Конечно, ее можно искусственно приписать одной из частей, т.е. считать, например, что кривая, разделяющая F_1 и F_2, принадлежит F_1, а множество F_2 в этой стороне «открыто», не имеет границы. Однако это условно: почему границу надо относить именно к F_1, а не к F_2? На самом деле все это несущественно, так как граница имеет площадь, равную нулю, и поэтому ее можно рассматривать отдельно, а можно приписывать к любой области. С этой оговоркой в аксиоме S_1 можно не требовать, чтобы фигуры F_1 и F_2 не имели общих точек, а фиксировать наличие общих точек (границы), так чтобы площадь этой общей части была равна нулю.

Все это связано с тем, что в математике принято считать, что точка, отрезок, кривая имеют нулевую площадь. Данное утверждение можно пояснить исходя из аксиом.

Пример

Пусть площадь некоторой точки A равна $S(A)$. Докажем, что $S(A) = 0$. Возьмем произвольное положительное число a и изобразим квадрат со стороной a так, чтобы точка A была внутри этого квадрата (например, в его центре; рис. 349). Обозначим построенный квадрат K_a. По аксиоме S_3, $S(K_a) = aa = a^2$. По теореме о монотонности площади, $S(A) = S \leq S(K_a) = a^2$. Итак, при любом положительном a получаем неравенство $0 \leq S \leq a^2$. Так как число a произвольно, то a^2 можно взять как угодно малым, поэтому S равно нулю, что и требовалось доказать.

Рис. 349

В дальнейшем при вычислении площадей мы уже не будем обращать внимание на то, включена граница в рассматриваемую фигуру или у всех рассматриваемых фигур площадь границы считается равной нулю.

Важные формулы планиметрии

Напомним различные формулы для вычисления площадей многоугольников:

1. *Площадь прямоугольного треугольника* (рис. 350):

$$S = \frac{1}{2}ab.$$

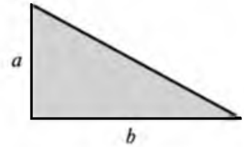


Рис. 350

Для доказательства формулы достроим треугольник до прямоугольника (рис. 351): По аксиоме S_4, площади полученных треугольников равны, а в сумме (по аксиоме S_2) они равны площади прямоугольника, которая, по аксиоме S_3, равна ab. Итак, площадь прямоугольного треугольника равна половине произведения длин катетов, т.е. $S = \frac{1}{2}ab$, что и требовалось доказать.

2. *Площадь произвольного треугольника* (рис. 352, *а*, *б*)

$$S = \frac{1}{2}ah.$$

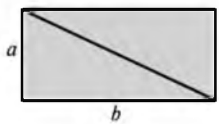


Рис. 351

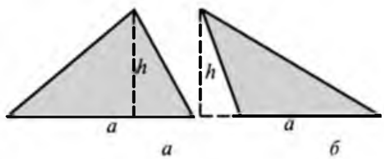


Рис. 352

Предлагаем вам самостоятельно провести доказательство этой формулы, используя рис. 353.

Напомним известные вам из планиметрии формулы для вычисления площадей многоугольников.

3. *Площадь параллелограмма* (рис. 354)

$$S = ah = ab\sin\alpha,\ S = \frac{1}{2}d_1d_2\sin\beta.$$

4. *Площадь треугольника* (рис. 355)

$$S = \frac{1}{2}bc\sin\alpha,\ S = rp,\ S = \frac{abc}{4R},\ S = \sqrt{p(p-a)(p-b)(p-c)},$$

где r — радиус вписанного круга; R — радиус описанного круга; p — полупериметр.

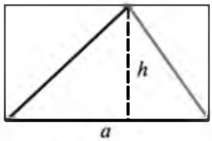


Рис. 353

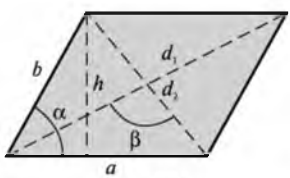


Рис. 354

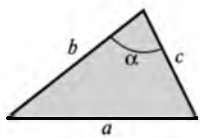


Рис. 355

5. *Площадь равнобедренного треугольника* (рис. 356)

$$S = \frac{1}{2}b^2\sin\alpha = \frac{1}{2}b^2\sin 2\beta,\ S = \frac{a^2}{4}\operatorname{tg}\beta.$$

6. *Площадь правильного треугольника* (рис. 357)

$$S = \frac{a^2\sqrt{3}}{4}, \; S = \frac{3R^2\sqrt{3}}{4},$$

где $a = R\sqrt{3}$; R — радиус описанного круга.

7. *Площадь правильного многоугольника с n сторонами* (рис. 358)

$$S_n = \frac{1}{2}nah, \; S_n = \frac{1}{2}nR^2\sin\frac{2\pi}{n},$$

где h — апофема; p — полупериметр.

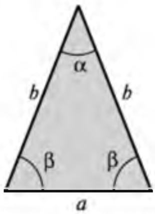


Рис. 356

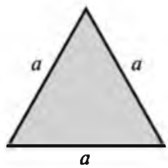


Рис. 357

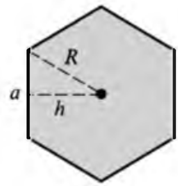


Рис. 358

При $n = 4$

$$S = a^2 = 2R^2;$$

при $n = 6$

$$S = \frac{3R^2\sqrt{3}}{2} = \frac{3a^2\sqrt{3}}{2}.$$

Выражение площади через интеграл

Приведенные ранее формулы позволили найти площади некоторых специальных многоугольников. Площадь произвольной фигуры может быть выражена через интеграл. Напомним, как это делается.

1. *Площадь подграфика*

$$S(F) = \int_a^b f(x)\mathrm{d}x$$

где $f(x) \geq 0$ (рис. 359).

Эта формула является определением интеграла.

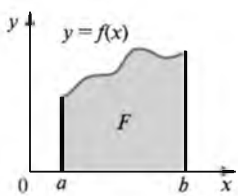


Рис. 359

2. *Площадь произвольной фигуры* (рис. 360). Введем систему координат так, чтобы границу фигуры F можно было разбить на графики двух функций (рис. 361).

Рис. 360

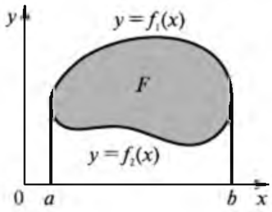


Рис. 361

Из аксиомы S_2 следует, что площадь подграфика функции f определяется суммой площадей фигуры F и подграфика функции f_2. Тогда

$$S(F)=\int_a^b f_1(x)\mathrm{d}x-\int_a^b f_2(x)\mathrm{d}x.$$

Схема вычисления площади с помощью интеграла изображена на рис. 362.

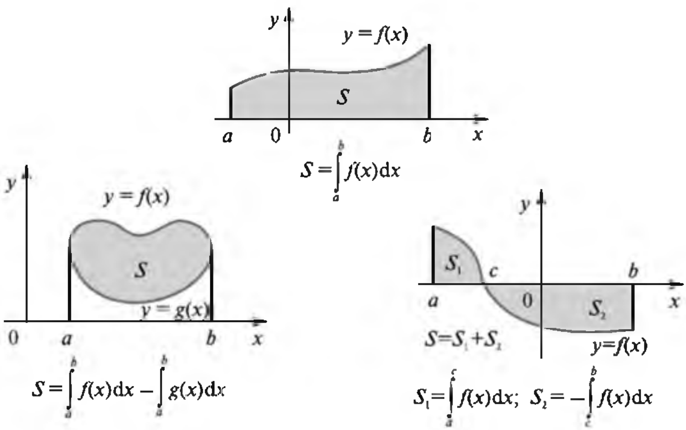


Рис. 362

Теперь определим площадь фигуры с использованием интеграла в том случае, когда надо самостоятельно вводить систему координат.

Пример

Вычислите площадь фигуры, изображенной на рис. 363, если известно, что закругление выполнено в форме квадратичной параболы.

Выберем оси координат так, как показано на рис. 364. Найдем уравнение квадратичной параболы. Так как ее вершина лежит на оси y, то $y = ax^2 + b$. Коэффициенты a и b вычислим, зная указанные на чертеже размеры: подставив в уравнение $x = 0$, найдем $b = 4$, подставив $y = 0$, получим $ax^2 + 4 = 0$, но из чертежа видно, что при $y = 0$ $x = \pm 8$, поэтому $a \cdot 8^2 + 4 = 0$, откуда $a = -\frac{1}{16}$. Итак, уравнение параболы $y = -\frac{1}{16}x^2 + 4$. Разобьем фигуру на треугольник и подграфик параболы. Площадь треугольника равна $\frac{1}{2} \cdot 16 \cdot 8 = 64$. Площадь подграфика вычисляем с помощью интеграла:

$$S = 2\int_0^8 \left(-\frac{1}{16}x^2 + 4\right)dx = 2\left(-\frac{1}{16}\frac{x^3}{3}\Big|_0^8 + 4 \cdot 8\right) =$$

$$= 2\left(-\frac{32}{3} + 32\right) = \frac{128}{3} \approx 42\frac{2}{3}.$$

Общая площадь: $64 + 42\frac{2}{3} = 106\frac{2}{3}$.

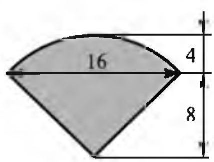


Рис. 363

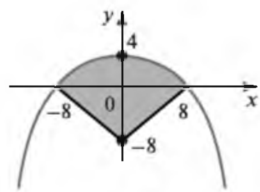


Рис. 364

Приближение площадей фигур многоугольниками

Со времен Архимеда (а возможно, и раньше) в геометрии использовалась следующая идея: для вычисления площади криволинейной фигуры ее заменяли приближенно многоугольником и пользовались формулами площади многоугольника.

Примеры использования этой идеи.

1. *Площадь круга.* Впишем в круг радиуса R и опишем около него правильные n-угольники. Пусть p_n и P_n — периметры n-угольников, s_n и S_n — их площади, h_n — апофема вписанного n-угольника. Получаем две последовательности $s_n = \frac{1}{2}h_n p_n$ и $S_n = \frac{1}{2}RP_n$, задающие приближения (с недостатком и с избытком) площади круга. Как, зная, что длина окружности равна $2\pi R$, получить площадь круга, равную πR^2? Можно проследить, к какому числу приближаются последовательности S_n и s_n.

В последовательности $S_n = \frac{1}{2}RP_n$ числа P_n приближаются к $2\pi R$, а все произведение — к числу $\frac{1}{2}R\cdot 2\pi R = \pi R^2$. Аналогично, в последовательности $s_n = \frac{1}{2}h_n p_n$ числа h_n приближаются к числу R, числа p_n — к $2\pi R$, а все произведение — к $\frac{1}{2}R\cdot 2\pi R = \pi R^2$.

Так как круг лежит между вписанным и описанным многоугольниками, то, по свойству монотонности площади, $s_n \le S \le S_n$. Ясно, что площадь круга отличается от площади правильного вписанного или описанного многоугольника меньше чем на $S_n - s_n$ (площадь заштрихованной на рис. 365 многоугольной фигуры).

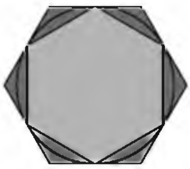


Рис. 365

2. *Приближенное вычисление интеграла.* Запишем формулу площади подграфика положительной функции:

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

Разобьем отрезок $[a, b]$ на n равных частей и заменим интегральной суммой — приближенной формулой прямоугольников. Пусть для простоты функция $f(x)$ возрастает на отрезке $[a, b]$. Замена подграфика внутренней ступенчатой фигурой означает замену интеграла S суммой

$$s_n = \frac{b-a}{n}\left(f(x_0) + f(x_1) + \ldots + f(x_{n-1})\right).$$

Внешняя ступенчатая фигура даст приближение интеграла S суммой $S_n = \frac{b-a}{n}\left(f(x_1) + f(x_2) + +\ldots + f(x_n)\right)$. Ясно, что истинное значение интеграла отличается от значения S_n или s_n не больше чем на $S_n - s_n$, т.е. не больше чем на площадь закрашенной серым цветом на рис. 366 многоугольной фигуры.

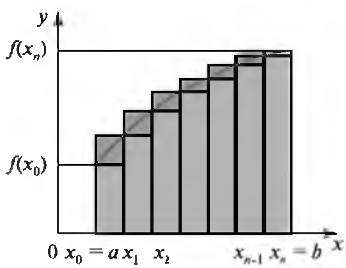


Рис. 366

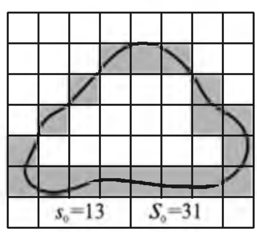


Рис. 367

3. *Применение палетки*. Нанесем фигуру, площадь которой мы измеряем, на лист бумаги, расчерченный на единичные квадраты. Подсчитаем количество квадратов, целиком содержащихся в данной фигуре. Обозначим это количество s_0, а S_0 — количество квадратов, имеющих с фигурой хотя бы одну общую точку.

Теперь нанесем сетку в 10 раз более мелкую. Площадь маленького квадрата равна 0,01. Пусть s_1 — количество этих квадратов, целиком лежащих внутри фигуры, S_1 — количество квадратов, имеющих с ней хотя бы одну общую точку. Числа $s_1 \cdot 10^{-2}$ и $S_1 \cdot 10^{-2}$ дают следующее приближение к площади фигуры. Продолжим процесс уменьшения размеров квадрата дальше. Каждый раз при этом мы заменяем фигуру некоторым многоугольником, площадь которого равна на n-м шаге $s_n \cdot 10^{-2n}$ или $S_n \cdot 10^{-2n}$. Эти числа дают приближенные значения площади фигуры.

Заменяя фигуру многоугольником, на n-м шаге мы совершаем ошибку, которая не больше чем $(S_n - s_n) \cdot 10^{-2n}$, где $S_n - s_n$ — количество квадратов, которые имеют хотя бы одну общую точку с фигурой, но не лежат целиком в ней, т.е. имеют точки, не принадлежащие фигуре. Они образуют многоугольную фигуру, покрывающую границу фигуры. Если граница фигуры имеет нулевую площадь, то число $(S_n - s_n) \cdot 10^{-2n}$ становится очень малым, т.е. погрешность измерения стремится к нулю.

Квадратную сетку, с помощью которой происходит измерение, называют палеткой (рис. 367).

Во всех трех рассмотренных случаях вычисления площади некоторой фигуры F мы поступали следующим образом: строили (способы построения были различными) две последовательности многоугольников: одну, составленную из многоугольников, целиком лежащих внутри фигуры F, другую — из многоугольников, покрывающих F. Измеряя площади этих многоугольников, мы получали две последовательности чисел, задающих приближения (с недостатком и с избытком) площади фигуры F. Разность этих приближений на n-м шаге дает оценку для ошибки измерения. Геометрически эта оценка совпадала с площадью многоугольной фигуры, которой можно покрыть границу фигуры F. Так как граница фигуры имеет нулевую площадь, то ошибку можно сделать сколь угодно малой.

Контрольные вопросы и задания

1. Перечислите аксиомы площади.
2. Какие вы знаете формулы для нахождения площади треугольника?

3. Что нужно знать, чтобы вычислить площадь фигуры с помощью интеграла?
4. Какие вы знаете приближенные методы вычисления площади?

§ 36. Объемы пространственных тел

Аксиомы объема

Изучение объемов мы проведем точно так же, как изучение площадей.

Аксиома V_1 (неотрицательность объема)
Объем V каждого измеряемого тела F есть неотрицательное число:

$$V(F) \geq 0.$$

Аксиома V_2 (аддитивность объема)
Если тело F разбито на две части F_1 и F_2, не имеющие общих точек, то объем тела есть сумма объемов тел F_1 и F_2:

$$V(F) = V(F_1) + V(F_2).$$

Аксиома V_3 (нормированность объема)
Объем прямого цилиндра равен произведению площади основания на высоту (рис. 368).

В частности, *объем единичного куба равен единице.*

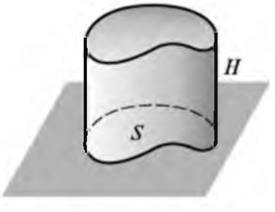


Рис. 368

Аксиома V_4 (инвариантность объема)
Равные тела имеют равные объемы.

Эта аксиома утверждает, что при перемещении тела его объем не изменяется.

Теорема о монотонности объема доказывается аналогично соответствующей теореме о площади.

Теорема. Объем части тела не больше объема всего тела (рис. 369): $V(F1) \leq V(F)$.

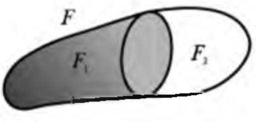


Рис. 369

Так же как и в случае площадей, можно доказать, что точка, кривая, поверхность имеют объем, равный нулю. Это позволяет при использовании аддитивности объема разбивать тело на две части некоторой поверхностью и не заботиться о том, какой части приписать эту поверхность — границу, так как она имеет нулевой объем.

Интегральная формула для объема

Вспомним, как мы ранее вычислили объем лимона. Лимон тогда резался на дольки с параллельными краями. Очень важно, что каждая долька настолько тонка, что ее можно считать цилиндром (с одинаковыми по площади основаниями). Правда, для разных долек эта площадь будет разной. Однако, зная, как меняется эта площадь в зависимости от того, где, в какой части лимона, вырезана долька, можно написать интегральную формулу для объема.

Рассмотрим тело F, выберем в пространстве некоторую ось x, и будем рассекать тело F плоскостями, перпендикулярными оси x.

Проекция тела F на ось x представляет некоторый отрезок, концы которого обозначим a и b. Плоскость сечения будем задавать точкой x на оси: $a \le x \le b$. Площадь сечения, проведенного через точку x, обозначим $S(x)$ (рис. 370). Таким образом, мы построили функцию — площадь переменного сечения $S = S(x)$, определенную при $x \in [a, b]$.

Из всего тела выделим маленькую часть, заключенную между сечениями, проведенными через точки x и $x + \mathrm{d}x$ (рис. 371). Эту маленькую часть тела можно приближенно считать цилиндром с основанием $S(x)$ и высотой $\mathrm{d}x$. Объем такого цилиндра, по аксиоме V_3, равен $S(x)\mathrm{d}x$. Следовательно, $\mathrm{d}V = S(x)\mathrm{d}x$, откуда

$$V = \int_a^b S(x)\,\mathrm{d}x\,.$$

Это и есть интегральная формула для объема.

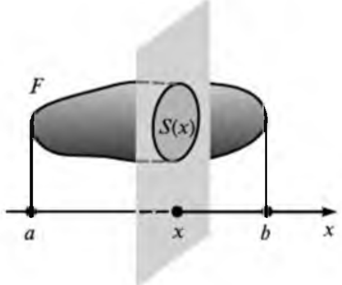


Рис. 370

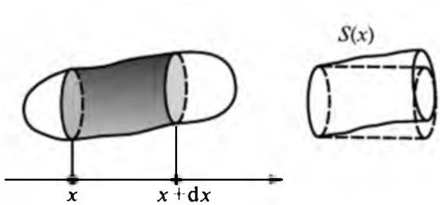


Рис. 371

Объем равен интегралу от площади переменного сечения.

При вычислении объема тела с помощью интегральной формулы направление, в котором проводятся плоские сечения, может быть произвольным. Его обычно выбирают так, чтобы формула для площади переменного сечения была возможно более простой. Мы убедимся в этом, вычисляя объемы пирамиды, конуса и шара.

Объем наклонного цилиндра

Рассмотрим наклонный цилиндр. Сечения этого цилиндра плоскостями, параллельными основаниям, имеют постоянную площадь S (рис. 372), поэтому ось x проведем перпендикулярно плоскостям оснований. Выберем начало O на оси x в точке пересечения оси с плоскостью, содержащей нижнее основание, а направление оси укажем снизу вверх. Тогда плоскость верхнего основания пересекает ось x в точке H, где H — высота цилиндра. Итак,

$$S(x) = S \text{ и } V = \int_0^H S\mathrm{d}x = S\int_0^H \mathrm{d}x = SH .$$

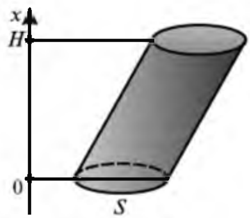


Рис. 372

Мы получили, что для объема наклонного цилиндра справедлива та же формула, что и для объема прямого цилиндра.

Объем пирамиды

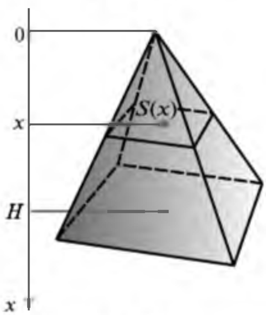


Рис. 373

Рассмотрим пирамиду (рис. 373). Ось x проведем перпендикулярно плоскости основания. Тогда в сечениях пирамиды плоскостями, параллельными основанию (а значит, перпендикулярными оси x), будут многоугольники, подобные основанию. Выберем начало O на оси x в точке пересечения оси с сечением, проходящим через вершину пирамиды, а направление — сверху вниз. Тогда плоскость основания пересекает ось в точке $x = H$, где H — высота пирамиды.

Вычислим площадь переменного сечения $S(x)$. В сечении получается многоугольник, подобный основанию. Коэффициент подобия равен отношению расстояний секущих плоскостей от вершины: $\frac{x}{H}$. Обозначим площадь основания S. По теореме об отношении площадей подобных многоугольников имеем

$$\frac{S(x)}{S}=\left(\frac{x}{H}\right)^2, \text{ т.е. } S(x)=\frac{S}{H^2}x^2.$$

Площадь переменного сечения представлена квадратичной функцией от x. Интеграл от нее вычисляется просто:

$$V=\int_0^H S(x)\mathrm{d}x=\frac{S}{H^2}\int_0^H x^2\mathrm{d}x=\frac{S}{H^2}\cdot\frac{x^3}{3}\bigg|_0^H=\frac{1}{3}SH.$$

Объем пирамиды равен одной трети произведения площади основания на высоту.

Объем конуса

Формула для объема конуса получается аналогично формуле для объема пирамиды (рис. 374):

$$S(x)=\frac{S}{H^2}x^2;\ V=\frac{1}{3}SH.$$

Объем любого конуса равен одной трети произведения площади основания на высоту.

В частности, объем кругового конуса с радиусом основания R и высотой H равен $V=\frac{1}{3}\pi R^2H$.

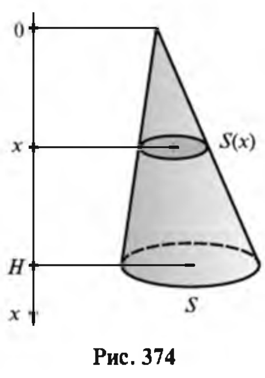


Рис. 374

Объем шара

В сечении шара любой плоскостью получается круг. Поэтому ось x можно выбрать произвольно, а начало O взять в точке пересечения плоскости, проходящей через центр шара (рис. 375, *а*) с осью x. Вычислим площадь круга, получающегося в сечении плоскостью, проходящей через точку x. По теореме Пифагора, $r^2 + x^2 = R^2$ (рис. 375, *б*), откуда

$$S(x) = \pi r^2 = \pi\,(R^2 - x^2).$$

Вследствие симметрии шара весь его объем V равен удвоенному объему его верхней половины, который получится при интегрировании переменной площади от 0 до R:

$$V = 2\int_0^R S(x)\mathrm{d}x = 2\int_0^R \pi(R^2 - x^2)\mathrm{d}x = 2\pi\left(R^2\int_0^R \mathrm{d}x - \int_0^R x^2\mathrm{d}x\right) =$$

$$= 2\pi\left(R^3 - \frac{x^3}{3}\bigg|_0^R\right) = 2\pi\left(R^3 - \frac{R^3}{3}\right) = 2\pi\frac{2R^3}{3} = \frac{4}{3}\pi R^3.$$

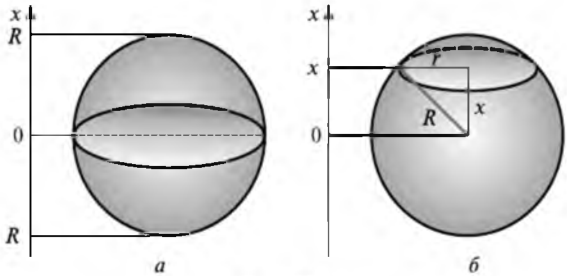


Рис. 375

Итак, объем шара радиуса R равен

$$V = \frac{4}{3}\pi R^3.$$

Объем тела вращения

Кроме шара существует и много других тел, в сечениях которых получаются круги. Представим себе тело, образованное вращением некоторой плоской фигуры относительно оси. Сечение этого тела, перпендикулярное оси вращения, — круг.

Пусть фигура представляет собой подграфик функции $y = f(x)$, заданной на отрезке $[a, b]$ $(f(x)) > 0)$. При вращении этого подграфика относительно оси x получим тело вращения (рис. 376).

Сечение, проведенное через точку x перпендикулярно оси, будет кругом радиуса $r = f(x)$. Площадь сечения $S = \pi r^2 = \pi f^2(x)$. По интегральной формуле, объем тела вращения

$$V = \pi\int_a^b f^2(x)\mathrm{d}x .$$

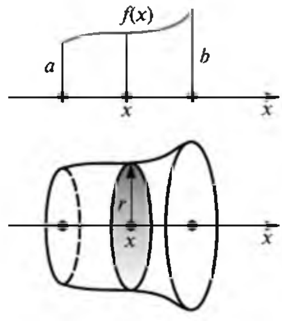


Рис. 376

Примеры

1. Найти объем тела, полученного при вращении отрезка гиперболы $y = \dfrac{1}{x}$, $x \in \left[\dfrac{1}{2}, 2\right]$, вокруг оси x:

$$V = \pi\int_{1/2}^{2}\frac{\mathrm{d}x}{x^2} = \pi\left(-\frac{1}{x}\right)\Bigg|_{1/2}^{2} = \pi\cdot\frac{3}{2} = \frac{3\pi}{2} .$$

2. Найти объемы тел, образованных вращением отрезка параболы $y = x^2$, $x \in [0, 1]$, вокруг оси x и вокруг оси y.
   Объем тела вращения относительно оси x вычисляется по формуле

$$V = \pi\int_0^1 (x^2)^2 dx = \pi\int_0^1 x^4 dx = \frac{\pi}{5} \approx 0{,}63 .$$

Для нахождения объема тела вращения относительно оси y зависимость $y(x) = x^2$ заменяется на зависимость $x(y) = \sqrt{y}$, где $y \in [0, 1]$:

$$V = \pi\int_0^1 (\sqrt{y})^2 \mathrm{d}y = \pi\int_0^1 y\,\mathrm{d}y = \frac{\pi}{6} \approx 1{,}05 .$$

Контрольные вопросы и задания

1. Перечислите аксиомы объема.
2. В чем состоит интегральная формула для объема?
3. Для каких тел вы знаете формулы для вычисления объема?
4. Как найти объем тела вращения?

§ 37. Площадь поверхности

Тела с плоскими развертками

Общая формула площади поверхности достаточно сложна. Мы ограничимся лишь простыми частными случаями.

Выведем формулу площади поверхности тела, когда его поверхность развертывается на плоскость. Простейшими примерами таких тел являются многогранники. Поверхность таких тел состоит из многоугольников, площади которых мы уже умеем вычислять, а площадь всей поверхности равна сумме площадей всех граней исходя из свойства аддитивности площадей.

Другим примером тел, допускающих развертку, являются цилиндры и конусы. Приведем формулы для площадей боковых поверхностей некоторых тел:

1) *правильная n-угольная призма со стороной основания a и высотой H*

$$S_{\text{бок}} = naH;$$

2) *правильная n-угольная пирамида со стороной основания a и апофемой h*

$$S_{\text{бок}} = \frac{1}{2}nah;$$

3) *прямой круговой цилиндр с радиусом основания R и высотой H*

$$S_{\text{бок}} = 2\pi RH;$$

4) *прямой круговой конус с радиусом R и образующей l*

$$S_{\text{бок}} = \pi Rl.$$

Выведем последнюю формулу.

Разверткой боковой поверхности прямого кругового конуса является круговой сектор, радиус которого равен длине образующей l, а длина дуги равна длине окружности основания конуса, т.е. $2\pi R$ (рис. 377). Используя формулу для площади кругового сектора, получаем

$$S_{\text{бок}} = \frac{1}{2}2\pi Rl = \pi Rl.$$

Нетрудно представить себе и другие тела, поверхности которых получены изгибанием плоских фигур. В этих случаях вычисление площади поверхности можно свести к вычислению площадей плоских фигур-разверток (рис. 378).

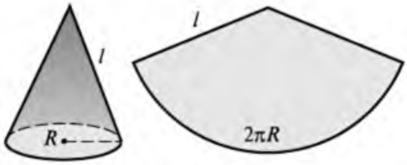


Рис. 377

Рис. 378

Поверхность шара

Далеко не всякое тело можно развернуть на плоскость. Простейшим примером может служить шар. Как же тогда измеряется площадь его поверхности?

Размер площади кривой поверхности можно определить по количеству краски, которую нужно потратить на равномерное покрытие поверхности. Предположим, что краска нанесена равномерно слоем толщины ε только на одну сторону поверхности. Масса краски пропорциональна объему ΔV тонкой пленки, который можно приближенно оценить как произведение искомой площади S на толщину ε слоя: $\Delta V = S\varepsilon$. Чем меньше ε, тем точнее формула. Можно вычислить площадь S приближенно как отношение $\dfrac{\Delta V}{\varepsilon}$ при малом ε, т.е. за площадь поверхности тела принять предел отношения объема слоя к его толщине при стремлении толщины этого слоя к нулю.

Применим указанную идею для вычисления площади поверхности шара. Пусть шар имеет радиус R. Тонкая пленка представит собой тело, заключенное между двумя концентрическими шарами радиусов R и $R + \mathrm{d}R$, где $\mathrm{d}R$ — толщина пленки (рис. 379).

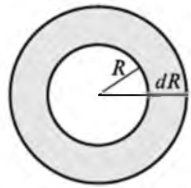


Рис. 379

Если $V(x)$ — объем шара радиуса x, то объем пленки равен $V(R + \mathrm{d}R) - V(R) = \Delta V$, т.е. приращению объема как функции радиуса.

Предел отношения $\dfrac{\Delta V}{\Delta R}$ есть производная объема шара по радиусу: $S(R) = \dfrac{\Delta V}{\Delta R}$. Зная формулу объема шара, дифференцированием получаем формулу для площади поверхности:

$$V(R) = \frac{4}{3}\pi R^3 \Rightarrow S(R) = \left(\frac{4}{3}\pi R^3\right)' = \frac{4}{3}\pi \cdot 3R^2 = 4\pi R^2 .$$

Итак, площадь поверхности шара радиуса R равна $4\pi R^2$.

Интересно заметить, что аналогично поверхности шара можно получить формулы для длин кривых, исходя из площадей. Так, длина окружности C получается дифференцированием формулы для площади круга:

$$S = \pi R^2 \Rightarrow C = S' = (\pi R^2)' = 2\pi R.$$

Контрольные вопросы и задания

1. Приведите примеры развертывающихся поверхностей.
2. Как вычислить площадь поверхности, если ее можно развернуть на плоскость?
3. Для каких тел вы знаете формулу вычисления площади поверхности?
4. Приведите примеры тел, поверхности которых не развертываются на плоскость.
5. Чему равна площадь поверхности шара?

Заключительная беседа

Составление дифференциального уравнения

Вернемся к задаче, в которой для вычисления произведенной работы мы ввели две переменные величины (см. главу 8): x — высоту от земли, на которую подняты камни, и A — работу, совершенную при возведении пирамиды до высоты x. Дальнейшие рассуждения дали соотношение между дифференциалами этих величин $\mathrm{d}A = F(x)\mathrm{d}x$, где $F(x) = \dfrac{\rho g a^2}{H^2} x(H - x)^2$. Полученное соотношение назовем дифференциальным уравнением для нахождения $A = A(x)$. Запишем его с помощью производной: $\dfrac{\mathrm{d}A}{\mathrm{d}x} = F(x)$ или $A' = F(x)$. Это уравнение очень простое — в нем производная неизвестной функции выражена через аргумент x. Отыскание самой функции A сводится к интегрированию:

$$A(x) = \int_0^x F(x)\mathrm{d}x\,.$$

Дифференциальные уравнения — это уравнения, связывающие неизвестную функцию и ее производные.

Многие физические законы имеют вид дифференциальных уравнений, т.е. соотношений между функциями и их производными. Задача интегрирования этих уравнений — важнейшая задача математики. Одни дифференциальные уравнения удается интегрировать в явном виде, т.е. записать искомую функцию в виде формул. Для решения других до сих пор не удается найти достаточно удобных формул. В этих случаях можно найти приближенные решения с помощью вычислительных машин. Мы не будем подробно изучать методы интегрирования дифференциальных уравнений, а только рассмотрим несколько примеров.

Примеры

1. *Уравнение механического движения*. Пусть материальная точка массы m движется под действием силы F по оси x. Обозначим t время ее движения, v — скорость, a — ускорение. Второй закон Ньютона $a = \dfrac{F}{m}$ примет вид дифференциального уравнения, если записать ускорение a как вторую производную: $a = \dfrac{d^2x}{dt^2} = x''$.

   Уравнение $mx'' = F$ называют *уравнением механического движения*, где $x = x(t)$ — неизвестная функция, m и F — известные величины. В зависимости от условий задачи по-разному и записываются различные дифференциальные уравнения.
   Например:

   а) сила постоянна ($F = \text{const}$). Уравнение движения $x'' = \dfrac{F}{m} = a$ (где a — const);

   б) сила периодически изменяется со временем, например по закону $F = F_0 \sin \omega t$. Уравнение движения $x'' = \dfrac{F_0}{m} \sin \omega t$;

   в) сила пропорциональна смещению (движение идеально упругой пружины): $F = -kx$ ($k > 0$), знак «–» указывает на то, что направление силы противоположно направлению смещения. Уравнение движения $x'' = -\dfrac{k}{m} x$;

   г) сила, обратно пропорциональная квадрату расстояния, $F = -\dfrac{km}{x^2}$ (свободный полет). Уравнение движения $x'' = -\dfrac{k}{x^2}$;

д) постоянная сила тяжести $F_1 = mg$ и сила трения $F_2 = -kx'$, пропорциональная скорости, действующие одновременно (падение с трением). Уравнение движения

$$x'' = g - \frac{k}{m}x'.$$

Во всех уравнениях движения вторая производная неизвестной функции x выражена через время t, положение x точки и ее скорость v. Такое уравнение называют *уравнением второго порядка*, так как в него входит вторая производная. Дифференциальное уравнение для работы по постройке пирамиды Хеопса является уравнением первого порядка, так как в нем имеется только первая производная.

2. *Радиоактивный распад*. Масса m радиоактивного вещества изменяется со временем: $m = m(t)$. Экспериментальные данные дают основание считать, что скорость изменения массы пропорциональна массе вещества в данный момент: $\frac{\mathrm{d}m}{\mathrm{d}t} = -km$, где k — коэффициент пропорциональности (знак «–» показывает, что масса вещества убывает).
3. *Народонаселение*. Представим число жителей страны в момент времени t как функцию $L = L(t)$. Допустим, что за единицу времени народонаселение увеличивается на определенный процент. Если в момент времени t число жителей равно $L(t)$, то за период времени $[t, t + \mathrm{d}t]$ появится примерно $kL(t)\,\mathrm{d}t$ новых жителей, т.е. $\Delta L \approx kL(t)\,\mathrm{d}t$. Хотя величина L принимает целые значения, обычно интересуются приближенными значениями L. Заменяя функцию L функцией, принимающей значения непрерывно и удовлетворяющей соотношению $\mathrm{d}L = kL\,\mathrm{d}t$, мы не сделаем большой ошибки. Таким образом, скорость роста функции L равна kL и она удовлетворяет дифференциальному уравнению $\frac{\mathrm{d}L}{\mathrm{d}t} = kL$.
4. *Электрическая цепь*. Пусть электрическая цепь состоит из последовательно соединенных резистора и конденсатора (рис. 380). В цепи произошло короткое замыкание и конденсатор, имевший начальный заряд, начинает разряжаться. Напряжение на конденсаторе в момент времени t обозначим $U = U(t)$. Заряд $q = q(t)$ связан с напряжением U формулой $q = CU$, где C — емкость конденсатора. Через сопротивление пойдет ток силой

$I = -\dfrac{U}{R}$, где R — сопротивление резистора, а знак «—» связан с направлением тока. Сила тока — это производная заряда по времени. Учитывая, что $I = -\dfrac{U}{R}$, а $q = CU$, получаем $C\dfrac{dU}{dt} = -\dfrac{U}{R}$ или $\dfrac{dU}{dt} = -\dfrac{1}{RC}U$.

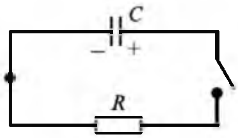


Рис. 380

Сформулируем вывод. Многие явления природы и техники описываются дифференциальными уравнениями, т.е. уравнениями, связывающими неизвестные величины и их производные. Вывод дифференциальных уравнений основан на знании законов изучаемых явлений.

Решение дифференциального уравнения

Дифференциальное и интегральное исчисления позволили записать на математическом языке в виде дифференциальных уравнений различные законы и явления. За 300 лет существования этого раздела математики появились многие тысячи дифференциальных уравнений. Однако замечательно то, что многие уравнения похожи друг на друга. Сравним, например, три уравнения, полученных в рассмотренных ранее примерах 2—4:

$$\frac{dm}{dt} = -km : \frac{dL}{dt} = kL ; \quad \frac{dU}{dt} = -\frac{1}{RC}U .$$

Все они имеют один и тот же вид — скорость изменения искомой функции пропорциональна значению этой функции. Решения всех трех уравнений сводятся к решению одного:

$$\frac{dx}{dt} = \alpha x .$$

(при разных значениях коэффициента пропорциональности α). Здесь мы наблюдаем удивительное могущество математики — совершенно разные процессы приводят к одной и той же математической модели. Ее исследование дает нам ответ, как в разобранных задачах, так и во многих других, которые приводят к такому же уравнению.

Математики научились объединять вместе похожие уравнения, классифицировать их. Так же как в случае простых алгебраических уравнений были найдены в свое время формулы для их корней, так и в случае некоторых стандартных дифференциальных уравнений были получены формулы для их решений.

Решение дифференциального уравнения — это функция, при подстановке которой в уравнение оно превращается в тождество. Так как операция дифференцирования выполняется просто, то всегда нетрудно проверить, является ли данная функция решением дифференциального уравнения или нет.

Примеры

1. Функция $x = \frac{at^2}{2} + v_0t + x_0$, где v_0, x_0 — произвольные числа, является решением уравнения $x'' = a$. Действительно, вычисляя производные, получаем $x' = at + v_0$, $x'' = a$.
2. Функция $x = -\frac{F_0}{m\omega^2} \sin \omega t + v_0t + x_0$ является решением уравнения $x'' = \frac{F}{m} \sin \omega t$.
3. Функция $x = A\sin\left(\sqrt{\frac{k}{m}}t + a\right)$, где A и a — произвольные числа, является решением уравнения $x'' = -\frac{k}{m}x$.
4. Функция $m = Ce^{-kt}$, где C — произвольная постоянная, является решением уравнения $m' = -km$.

Однако не для всех уравнений решения записываются так просто. Уже для уравнения свободного космического полета написать формулу решения довольно трудно. Часто удается исследовать решение дифференциального уравнения, не находя самого решения.

Уравнение показательного роста

Известно много процессов, в которых скорость изменения какой-либо величины пропорциональна ее значению. К числу таких процессов относят радиоактивный распад, остывание тела, разряд емкости через сопротивление, изменение народонаселения и др.

Эти процессы описываются дифференциальным уравнением первого порядка $x' = ax$ при различных значениях коэффициента a.

Как же решить уравнение $x' = ax$?

Мы знаем, что показательная функция обладает тем свойством, что ее производная пропорциональна ей самой. Таким образом, функция $x' = e^{at}$ является одним из решений уравнения $x' = ax$. Как найти все решения? Пусть z — произвольное решение, т.е. $z' = az$. Запишем $z = ye^{at}$, где y — новая неизвестная функция. Тогда

$$(ye^{at})' = a(ye^{at}).$$

Вычислим производную левой части:

$$(ye^{at})' = y'e^{at} + aye^{at}.$$

Приравнивая правые части уравнений, находим $e^{at}y' = 0$, т.е. $y' = 0$. Таким образом, производная функция равна нулю и $y = C$. Итак, любое решение уравнения $x' = ax$ имеет вид $x = Ce^{at}$, где C — константа.

Получаем решения для рассмотренных уравнений.

Значение константы C определяем, зная начальное значение искомой величины. Если в качестве начального момента времени взято $t = 0$, то значение C как раз и равно значению искомой величины при $t = 0$. Так, если начальное напряжение на конденсаторе U_0, то конденсатор разряжается по экспоненциальному закону: $U = U_0 e^{-\frac{t}{RC}}$. При больших t значения U приближаются к нулю. Графики показательных функций при разных a приведены на рис. 381.

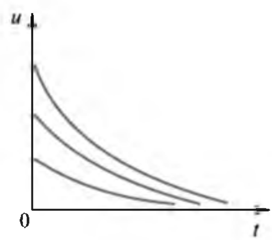


Рис. 381

Уравнение гармонических колебаний

Напомним уравнение движения материальной точки, прикрепленной к концу упругой пружины: $mx'' = -kx$. Оказывается, есть много задач, приводящих к аналогичным уравнениям второго порядка.

Пример

Рассмотрим электрическую цепь, состоящую из последовательно соединенных конденсатора емкостью C и катушки индуктивностью L (рис. 382). На конденсаторе есть начальное напряжение. В начальный момент времени цепь замкнули накоротко и через катушку пошел ток. Обозначим $U = U(t)$ напряжение на конденсаторе в момент времени t. На-

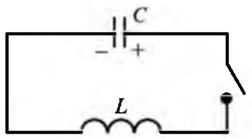


Рис. 382

пряжение $U(t)$ на катушке (при выбранном направлении тока) пропорционально скорости изменения силы тока I, проходящего через катушку: $U = -L\frac{dI}{dt}$. Подставляя в это выражение $I = \frac{dq}{dt}$, где $q = CU$, получаем:

$$U = -L\frac{dI}{dt} = -L\frac{d^2q}{dt^2} = -LC\frac{d^2U}{dt^2}.$$

Окончательно приходим к уравнению второго порядка относительно напряжения:

$$\frac{d^2U}{dt^2} = \frac{1}{LC}U.$$

Итак, две разные на первый взгляд задачи физики — колебание упругой пружины и разряд конденсатора через катушку — привели к одному и тому же дифференциальному уравнению второго порядка, только записанному в разных обозначениях:

Разберем первое из них, обозначив $-\omega^2$ константу, стоящую перед искомой функцией $x'' = -\omega^2 x$. Это уравнение называется уравнением свободных гармонических колебаний.

Проверим, что функция $x = A\cos(\omega t + a)$, где A и a — константы, является решением уравнения $x'' = -\omega^2 x$. Действительно,

$$x' = -A\,\omega\sin(\omega t + a),\ x'' = -A\,\omega^2\cos(\omega t + a) = -\omega^2 x.$$

Оказывается, что, изменяя A и a, мы получаем все решения уравнения гармонических колебаний. Константы A и a имеют наглядный смысл: A — это амплитуда колебаний, a — начальная фаза. Значения A и a находятся из начальных условий — значений x и x' в начальный момент времени. Графики гармонических колебаний при различных значениях A, ω и a приведены на рис. 383.

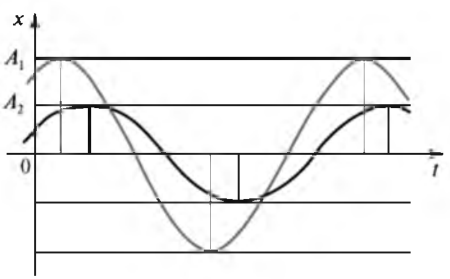


Рис. 383

Уравнение гармонических колебаний получено нами при предположениях, которые реально не выполняются. Так, при описании колебания пружины надо учитывать трение, а при описании разряда конденсатора — внутреннее сопротивление. При этом в уравнении колебаний появляется дополнительное слагаемое, зависящее от первой производной (скорости).

Пример

Уравнение для силы тока I в цепи (рис. 384).

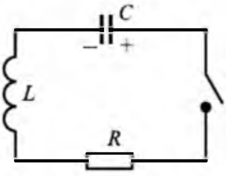


Рис. 384

$$L\frac{d^2I}{dt^2}+R\frac{dI}{dt}+\frac{1}{C}I=0\ .$$

Решения такого уравнения зависят от соотношения между параметрами L, R, C.

Свободные колебания наблюдаются в электрической цепи, если в ней произошло короткое замыкание и ток идет только за счет начального заряда конденсатора. Если в цепи подключить источник ЭДС, то колебания тока описываются новым уравнением, которое нетрудно вывести. Решения такого уравнения характеризуют вынужденные колебания, зависящие от подаваемого напряжения.

Мы рассмотрели некоторые примеры дифференциальных уравнений. При этом мы не ставили задачи научиться решать эти уравнения: это предмет специального раздела математики, теории дифференциальных уравнений, которая изучается в высшей школе. Важно понять, что с помощью основных операций анализа (дифференцирования и интегрирования) можно строить математические модели (дифференциальные уравнения) достаточно сложных и важных процессов. Полезно иметь в виду и то, что разные задачи могут приводить к одной и той же модели, что делает наиболее часто встречающиеся уравнения особенно важными. К их числу относят уравнение показательного роста, уравнения свободных и вынужденных колебаний.

Контрольные вопросы и задания

1. Что такое дифференциальное уравнение?
2. Как записать второй закон Ньютона в виде дифференциального уравнения?
3. Приведите примеры дифференциальных уравнений механического движения.
4. Каким дифференциальным уравнением описывается радиоактивный распад?

5. Какому дифференциальному уравнению в первом приближении удовлетворяет функция, описывающая рост народонаселения?
6. Напишите дифференциальное уравнение свободных гармонических колебаний.
7. Приведите примеры разных процессов, описываемых одним и тем же дифференциальным уравнением.
8. Какой вид имеют решения уравнения показательного роста?
9. Чему равна частота колебания, описываемого уравнением $x'' + kx = 0$?
10. Как определяются значения констант, появляющихся при решении дифференциальных уравнений?

ГЛАВА 9

КОМБИНАТОРИКА

§ 38. Правило произведения

Перебор вариантов

С перебором возможных вариантов человек сталкивается в своей жизни очень часто: выбор способа передвижения к месту работы или учебы, распределение порядка выполнения намеченных дел, планирование свободного времени и др. А вспомните игру в шахматы или любую другую игру. Делая очередной ход, вы анализируете множество различных продолжений и останавливаете свой выбор на том, которое вам кажется наиболее выгодным.

Общие законы комбинирования вариантов, формулы для подсчета их числа, правила их перебора составляют часть математики, которая называется *комбинаторикой*. Толчком к развитию этой древней науки явилось появление новой вычислительной техники, которая предоставила возможность осуществлять переборы миллионов вариантов. С другой стороны, многие важные прикладные задачи при решении их простым перебором вариантов приводят к такому объему вычислений, которые в принципе нельзя реализовать на вычислительной машине. Это заставило искать новые алгоритмы целенаправленного перебора вариантов, совершенствовать методы подсчета или оценки их количества.

Мы ознакомимся с простейшими комбинаторными задачами, в которых будет идти речь в основном о подсчете количества различных комбинаций с теми или иными свойствами.

Составление таблиц

Представим себе группу из пяти мальчиков (Боря, Володя, Гриша, Дима и Женя) и четырех девочек (Аня, Ира, Оля и Лена). Как удобнее всего перебрать все пары, состоящие из одного мальчика и одной девочки? Естественно составить таблицу (табл. 9), в строки которой выписать имена мальчиков, а в столбцы — имена девочек. Каждая клетка этой таблицы соответствует одной паре «мальчик — девочка», причем все возможные пары будут перебраны. (В таблице выписаны только первые буквы имен.)

Таблица 9

Варианты перебора

Мальчики	Девочки			
	А	И	О	Л
Б	БА	БИ	БО	БЛ
В	ВА	ВИ	ВО	ВЛ
Г	ГА	ГИ	ГО	ГЛ
Д	ДА	ДИ	ДО	ДЛ
Ж	ЖА	ЖИ	ЖО	ЖЛ

Прямоугольные таблицы используются очень часто. Посмотрите на ваши учебные кабинеты. Вы увидите периодическую таблицу элементов Д.И. Менделеева, таблицу растворимости соединений, таблицу значений тригонометрических функций и др. Таблица позволяет удобно расположить пары элементов двух множеств. При этом количество таких пар подсчитать легко. В нашей таблице 20 клеток, что соответствует тому, что из пяти мужских имен и четырех женских можно составить $5 \cdot 4 = 20$ пар имен.

Полученный вывод можно записать в общем виде.

Правило произведения

Пусть даны два конечных множества A и B. Пусть множество A содержит m элементов, а множество B – n элементов. Число различных пар, которые можно составить, взяв по одному элементу из каждого множества, равно произведению mn.

Для доказательства этого правила осуществим перебор всех пар, составим таблицу, в строки которой поместим элементы множества A (ихm штук), а в столбцы — элементы множества B (их n штук). Клетки этой таблицы соответствуют всем искомым парам. Число клеток в таблице равно mn. Правило доказано.

Пример

На пяти перекрестках зажжено пять трсхсекционных светофоров (каждый светофор находится в одном из трех положений —

зажжен красный, зеленый или желтый свет). Сколькими способами это можно сделать, если считать, что светофоры работают друг от друга независимо?

Первый светофор можно зажечь тремя способами, второй тоже тремя, третий — тремя и т.д. Общее число способов равно $3 \cdot 3 \cdot 3 \cdot 3 \cdot 3 = 3^5 = 243$.

Выбор представителей

Рассмотрим задачу: «В училище есть три группы первого курса. В первой — 20 человек, во второй и третьей — по 25. Нужно выбрать по одному представителю от каждой группы. Каким числом способов это можно сделать?».

Если бы групп было две, то мы сделали бы это, как в рассмотренной задаче образования пар из элементов двух множеств. Здесь необходимо получать тройки представителей по одному человеку от каждой группы. Сделать таблицу теперь труднее. Конечно, можно мысленно представить себе трехмерную таблицу в пространстве в виде параллелепипеда, по каждому из трех измерений которого выписаны фамилии участников соответствующей группы. Однако на практике реализовать такую таблицу нелегко. Ответить же на вопрос: «Сколько клеток (кубиков) в такой таблице?» — можно: число равно произведению $20 \cdot 25 \cdot 25 = 12\,500$.

Можно рассуждать иначе. Сначала выберем одного представителя от первой группы. Это можно сделать 20 способами (в первой группе 20 человек). Затем выберем одного представителя от второй группы и составим пары (представитель первой группы, представитель второй группы). По правилу произведения таких пар будет $20 \cdot 25 = 500$. Затем, к каждой паре подсоединяя представителя из третьей группы, мы как бы снова составляем «пару», первый элемент которой — одна из полученных 500 пар, а второй элемент — один представитель третьей группы. Общее количество таких «пар» равно $500 \cdot 25 = 12\,500$.

Таким же рассуждением можно обобщить правило произведения для любого количества множеств.

Число слов

Многие задачи комбинаторики удобно рассматривать на следующей «языковой» модели. Исходное множество — это множество букв. Будем называть его *алфавитом* (можно, например, взять русский алфавит из 33 букв или латинский алфавит из 26 букв). Назовем *словом* произвольную (конечную) последовательность букв. Например, ВЕРА, АРБА, КЛАН, БРКС, ТТТТ — четырехбуквенные «слова», составленные из русского алфавита. Другой пример, номер автомашины состоит из трехбуквенного «слова» и «слова», составленного из «цифрового алфавита», скажем «ММЕ 27—31».

Пусть в алфавите 20 букв. Сколько можно составить двухбуквенных слов из этого алфавита? Двухбуквенное слово — это пара, на каждом месте которой стоит только одна из букв данного алфавита. Число таких пар можно подсчитать с помощью таблицы. Оно равно произведению $20 \cdot 20 = 400$. Некоторое отличие здесь от рассматривавшейся ранее задачи о числе пар состоит в том, что элементы пары теперь берутся из одного и того же множества («алфавита»), хотя эти элементы берутся независимо один от другого (вторая буква может совпадать с первой). Поэтому общее число пар равно снова произведению (совпадающих на этот раз) множителей: $20 \cdot 20 = 400$.

А как подсчитать число трехбуквенных слов? Трехбуквенные слова — это тройки, составленные из букв данного алфавита. Если в алфавите 20 букв, то число трехбуквенных слов равно $20 \cdot 20 \cdot 20 = 20^3 = 8000$. Для получения этого результата надо представить, как мы пишем трехбуквенное слово. Сначала пишем первую букву (20 вариантов); затем приписываем вторую букву (число вариантов возрастает в 20 раз, так как к каждой букве, написанной первой, мы можем приписать независимо любую из 20 букв), получится 20^2 двухбуквенных слов. Затем приписываем третью букву, число слов возрастает снова в 20 раз и получается $20^2 \cdot 20 = 20^3 = 8000$.

Таким образом, из алфавита в 20 букв можно составить 20 однобуквенных слов, 20^2 двухбуквенных слов, 20^3 трехбуквенных. Ясно, что четырехбуквенных слов в 20 раз больше, чем трехбуквенных, т.е. их число равно 20^4. В общем виде число k-буквенных слов в этом алфавите равно 20^k.

Если в алфавите число букв равно n, то в предыдущей формуле для числа k-буквенных слов надо заменить 20 на n и получить следующую формулу.

Число k-буквенных слов в алфавите из n букв равно nk.

Чтобы в этой формуле не перепутать местами буквы n и k, полезно каждый раз повторять рассуждения, сделанные при ее выводе.

Задача

На рояле 88 клавиш. Сколькими способами можно последовательно извлечь 6 звуков?

Первый звук можно извлечь 88 способами (нажать любую клавишу). При добавлении второго звука число вариантов увеличивается в 88 раз, т.е. два звука последовательно можно извлечь 88^2 способами, три звука — 88^3, четыре — 88^4 способами и т.д.

Ответ: 88^6.

При решении этой задачи мы не сводили последовательности звуков к словам в некотором алфавите. Это можно было сделать, применив обычную нотную запись звука. Знакомы ли вы с таким алфавитом? Сколько разных букв вы в нем знаете?

Составление расписания

Пусть мы составляем последовательность элементов данного множества, например, выписываем в виде слова буквы данного алфавита, составляем числа из цифр, мелодии из нот и др. Так подсчитывают число получаемых комбинаций в том случае, когда члены этой последовательности берутся независимо друг от друга. Однако часто встречаются задачи на подсчет числа комбинаций, в которых выбираемые элементы связаны друг с другом какими-то ограничениями, зависимостями. Важным примером такой задачи является задача о составлении расписания.

Задача

В классе 20 человек. Сколькими способами можно составить расписание дежурства на 6 дней недели так, чтобы каждый день дежурил один человек и никто не дежурил дважды?

Будем составлять список дежурных. На понедельник можно назначить дежурным любого из 20 человек. На вторник выбрать дежурным любого человека из 19 оставшихся. Таким образом, на 2 дня можно выбрать дежурных числом способов, равным $20 \cdot 19 = 380$.

Обратите внимание на последнее место. Если бы мы выбирали пару дежурных независимо друг от друга, то получили бы $20 \cdot 20 = 400$ вариантов. Выбор первого дежурного накладывает ограничение — его нельзя выбирать во второй день. (Подумайте, каким клеткам в таблице пар соответствуют возможные пары дежурных на два дня.)

Продолжим рассуждение. Выбрав дежурных на первые 2 дня $20 \cdot 19$ способами, на третий можно выбрать любого из оставшихся 18 человек, т.е. число списков дежурных увеличится в 18 раз. Продолжая выбор дежурных, мы получим, что на 6 дней можно выбрать дежурных числом способов, равным $20 \cdot 19 \cdot 18 \cdot 17 \cdot 16 \cdot 15$ (6 множителей). Задача решена.

Составление расписания дежурств можно иначе себе представить так: имеется 20 человек и 6 мест (дни недели). Нужно разместить на этих местах по одному человеку из данных 20 человек. Такие комбинации так часто и называют — *размещения*.

В общем виде: пусть у нас есть n человек и k мест ($k \leq n$). Каким числом способов можно разместить на этих местах (по одному на каждом месте) людей из данного множества?

Рассуждаем так же, как при составлении расписания дежурств: на первое место — любой из n человек, на второе — любой из ($n-1$) оставшихся, на третье — любой из ($n-2$) оставшихся и т.д. Число вариантов будет перемножаться (см. ранее, правило произведения). Общее число множителей равно k. Как записать произведение k последовательных целых чисел в убывающем порядке, самое большое из которых равно n? Это можно сделать так:

$$n(n-1)(n-2) \dots (n-k+1).$$

Действительно, если первое число равно n, второе $n-1$, третье $n-2$, то k-е число равно $n-(k-1)=(n-k+1)$. При определенных n и k формулу для числа размещений писать нетрудно — мы уже это делали, например, в ответе на задачу о расписании дежурств: $20 \cdot 19 \cdot 18 \cdot 17 \cdot 16 \cdot 15$.

Таким образом, мы доказали следующую формулу для числа размещений n объектов на k местах, которое обозначается A_n^k:

$$A_n^k = n(n-1) \dots (n-k+1).$$

Сделаем то же самое замечание, что и после вывода формулы для числа слов: не надо запоминать формулу для числа размещений. При решении задач, в которых приходится размещать по местам какие-то элементы, надо повторить те же рассуждения, которые были проделаны при решении задачи о составлении расписания.

Перестановки

Вернемся к задаче о составлении расписания. Представим, что нужно составить расписание дежурств 20 человек не на 6 дней, а на 20.

Каждый человек будет дежурить один раз. Рассуждая, как и раньше, мы получим формулу для числа таких расписаний: на первый день мы назначим любого из 20, на второй — любого из оставшихся 19 и т.д. В ответе получится число, составленное из 20 множителей, из которых самый большой 20: $20 \cdot 19 \cdot 18 \ldots$ Последнее число равно единице, т.е. мы получим произведение всех целых чисел от 1 до 20.

В общем виде произведение целых чисел от 1 до n включительно обозначается $n!$ (читается: «n-факториал».) Таким образом, $1! = 1$, $2! = 2$, $3! = 1 \cdot 2 \cdot 3 = 6$, $4! = 1 \cdot 2 \cdot 3 \cdot 4 = 24$, $5! = 1 \cdot 2 \cdot 3 \cdot 4 \cdot 5 = 120$ и т.д. Заметим, что факториалы можно получить последовательно: $6! = 5! \cdot 6 = 120 \cdot 6 = 720$. В общем виде $n! = n(n-1)!$

Итак, в рассмотренной задаче число мест, на которых мы должны разместить людей, равно числу этих людей. Такая ситуация встречается часто — рассадить n человек на n мест, или, что то же самое, приписать каждому человеку номер. Говорят еще так. Мы хотим упорядочить данное множество людей, расставить их по местам. Так как расположение элементов множества в определенном порядке встречается часто, то для таких расположений в математике употребляется специальное слово — *перестановка*.

Перестановкой элементов данного множества называется расположение их в определенном порядке.

Примеры

1. Заданное множество состоит из букв A и B. Перестановок всего две: AB и BA.
2. Заданное множество состоит из трех цифр: 1, 2 и 3. Перестановок шесть: 123, 132, 213, 231, 312, 321. Число перестановок из n элементов обозначается P_n. Мы вывели формулу

$$P_n = n\,(n-1) \cdot \ldots \cdot 2 \cdot 1 = n!$$

С увеличением n число перестановок растет очень быстро. Так, $n!$ при $n = 10$ равно 3 628 800, а при $n = 100$ оно имеет порядок 10^{158}.

Контрольные вопросы

1. В чем состоит правило произведения?
2. Как подсчитать число слов в алфавите, состоящем из n букв?
3. Что такое размещение?
4. Что такое перестановка?
5. Как вычислить число перестановок n предметов?

§ 39. Биномиальные коэффициенты

Число сочетаний

Решим задачу: «В группе 20 человек. Сколькими способами можно из них выбрать двух дежурных?». Вопрос выбора пары человек из данной группы мы уже обсуждали в предыдущем параграфе. Возможны три постановки задачи. Можно фиксировать порядок, в котором будут дежурить два человека. Именно эту задачу мы решали раньше, назначая по одному дежурному на каждый день. При этом есть два варианта: разрешать одному и тому же человеку дежурить оба дня (*ответ*: $20^2 = 400$ способов) или не разрешать (*ответ*: $20 \cdot 19 = 380$ способов). Можно рассматривать задачу выбора двух дежурных, не фиксируя их порядка, т.е. назначая их как бы на один день. Число способов при этом уменьшится.

Как же подсчитать число различных пар? Число пар различных людей с учетом порядка (как говорят, упорядоченных пар) равно, как мы знаем, $20 \cdot 19 = 380$. При этом, например, пары *АБ* и *БА* считаются различными. Если мы их будем считать одинаковыми, т.е. не будем интересоваться порядком, то число пар сократится вдвое: $\frac{20 \cdot 19}{2} = 190$.

Проведенные рассуждения можно наглядно изобразить с помощью таблицы. Составляя квадратную таблицу всех пар, которые можно составить из группы в 20 человек, получим всего клеток $20^2 = 400$. Это соответствует числу всех упорядоченных пар (в том числе таких, в которых элементы повторяются). Если мы вычтем диагональные клетки (их 20 шт.), то останется 380 клеток, что соответствует числу упорядоченных пар с неповторяющимися элементами. Если теперь отождествим клетки, симметричные относительно диагонали, то число пар уменьшится вдвое и их станет 190.

Наборы элементов данного множества без учета их порядка называют сочетаниями (их можно было бы назвать подмножествами данного множества).

Ранее мы решили задачу подсчета числа сочетаний из 20 элементов по 2. В общем случае число сочетаний из n элементов по k обозначается так: C_n^k (читается: «цэ из n по k»). Если заменить в разобранной задаче число 20 на n, то таким же рассуждением получим формулу

$$C_n^2 = \frac{n(n-1)}{2}.$$

Напомним, что $n(n-1)$ — это число упорядоченных пар с разными элементами (число размещений). Число же сочетаний вдвое меньше.

Выведем теперь общую формулу для числа сочетаний C_n^k, при любом k. Начнем опять с примера.

Задача

Каким числом способов можно выбрать из группы в 20 человек команду из 3 человек?

Сначала выберем упорядоченные тройки различных людей. Число таких троек равно $20 \cdot 19 \cdot 18$. Рассмотрим одну тройку, которую обозначим A, B, C. Ясно, что, переставляя этих людей, получим разные упорядоченные тройки, но одну и ту же команду, если нам не важен порядок элементов A, B и C. Значит, число команд в 6 раз меньше числа упорядоченных троек. Итак, искомое число команд или сочетаний из 20 по 3 равно $\frac{20 \cdot 19 \cdot 18}{6} =$ $= 1140$. Итак, $C_{20}^3 = \frac{20 \cdot 19 \cdot 18}{6}$.

Для любого n получим

$$C_n^3 = \frac{n(n-1)(n-2)}{6} = \frac{n(n-1)(n-2)}{3!}.$$

Если находить число команд не по три, а по k человек, то надо разделить число размещений (упорядоченных наборов по 6 человек) не на $3!$, а на $k!$:

$$C_n^k = \frac{n(n-1)...(n-k+1)}{k!}.$$

Эта формула в основном используется для небольших значений k. Отметим, что

$$C_n^1 = n, \quad C_n^2 = \frac{n(n-1)}{2}, \quad C_n^3 = \frac{n(n-1)(n-2)}{6}.$$

Решение смешанных комбинаторных задач

Приведем несколько примеров решения комбинаторных задач.

Задачи

1. Сколько различных результатов может быть в тираже «Спортлото 5 из 36»?

Из 36 номеров нужно выбрать 5 (разумеется, без учета порядка). Число таких выборов равно C_{36}^5, или

$$C_{36}^5 = \frac{36 \cdot 35 \cdot 34 \cdot 33 \cdot 32}{5!} = 376\,992.$$

2. Вы заполнили карточку «Спортлото 5 из 36». Сколько возможно различных результатов тиражей, в которых будет ровно 3 из отмеченных вами номеров?
Сначала выберем из отмеченных пяти номеров те три, которые попадут в тираж. Это можно сделать числом способов, равным

$$C_5^3 = \frac{5 \cdot 4 \cdot 3}{6} = 10.$$

Оставшиеся в тираже два номера могут быть любыми из числа номеров, вами не отмеченных (их осталось 31). Эти два номера можно выбрать числом способов, равным

$$C_{31}^2 = \frac{31 \cdot 30}{2} = 465.$$

Общее число способов находят, комбинируя все выборы первых трех и последних двух номеров по правилу произведения. *Ответ*: $465 \cdot 10 = 4650$.
Заметим, что данное число от общего числа вариантов составляет примерно 0,012, т.е. 1,2%.

3. В группе 20 человек. Каким числом способов можно выбрать из нее команду из трех человек и, кроме того, одного капитана?
Эту задачу можно решать различными способами.
Способ 1. Сначала выберем капитана (20 способов), а затем из оставшихся 19 человек выберем команду из трех человек ($C_{19}^3 = \frac{19 \cdot 18 \cdot 17}{6} = 969$ способов). Составляя комбинации по правилу произведения, получим *ответ*: $969 \cdot 20 = 19\,380$ способов.
Способ 2. Сначала выберем команду ($C_{20}^3 = \frac{20 \cdot 19 \cdot 18}{6} = 1140$ способов). Затем из оставшихся 17 человек выберем капитана (17 способов). Комбинируя варианты, получим *ответ*: $1140 \cdot 17 = 19\,380$ способов.
Способ 3. Сначала выберем четырех человек ($C_{20}^4 = \frac{20 \cdot 19 \cdot 18 \cdot 17}{4!} = 4845$ способов), а затем из четырех выбранных человек одного назначим капитаном (4 способа). *Ответ*: $4845 \cdot 4 = 19\,380$ способов.

В комбинаторике часто пользуются тем, что одну и ту же задачу можно решить по-разному: сравнивая ответы, полученные различными способами, можно получить интересные тождества. Так, в нашей задаче мы получили равенства $20 \cdot C_{19}^{3} = 17 \cdot C_{20}^{3} = 4 \cdot C_{20}^{4}$. (Попробуйте получить обобщения этих соотношений, обозначив количество человек в группе через n, а количество человек в команде — через k.)

Бином Ньютона

Коэффициенты C_n^k для подсчета числа сочетаний называются *биномиальными коэффициентами*, потому что они используются в формуле бинома Ньютона. Бином Ньютона служит для возведения в степень суммы двух слагаемых (бином — это двучлен).

Формулы для квадрата и куба суммы хорошо известны:

$$(a + b)^2 = a^2 + 2ab + b^2,\ (a + b)^3 = a^3 + 3a^2b + 3ab^2 + b^3.$$

Пример

Возвести в степень $(1 + x)^4$:

$$(1 + x)^4 = 1 + C_4^1x + C_4^2x^2 + C_4^3x^3 + C_4^4x^4 = 1 + 4x + 6x^2 + 4x^3 + x^4.$$

Аналогично записывается формула в общем виде:

$$(1 + x)^n = 1 + C_n^1x + C_n^2x^2 + \ldots + C_n^{n-1}x^{n-1} + x^n.$$

Свойства биномиальных коэффициентов

Выпишем биномиальные коэффициенты C_n^k в виде треугольной таблицы. Для удобства допустим, что k принимает нулевое значение, полагая $C_n^0 = 1$ (это соглашение можно оправдать так: из k элементов 0 элементов можно выбрать одним способом — ничего не взяв). Таблицу построим следующим образом:

$$
\begin{array}{ccccccc}
 & & & C_0^0 & & & \\
 & & C_1^0 & & C_1^0 & & \\
 & C_2^0 & & C_2^1 & & C_2^0 & \\
C_3^0 & & C_3^1 & & C_3^2 & & C_3^0
\end{array}
$$

Подставляя вместо C_n^k их числовые значения, получим таблицу:

$$\begin{array}{ccccccccccc} &&&&&1&&&&& \\ &&&&1&&1&&&& \\ &&&1&&2&&1&&& \\ &&1&&3&&3&&1&& \\ &1&&4&&6&&4&&1& \\ 1&&5&&10&&10&&5&&1 \\ \ldots&\ldots&\ldots&\ldots&\ldots&\ldots&\ldots&\ldots&\ldots&\ldots&\ldots \end{array}$$

В строчках таблицы стоят коэффициенты, участвующие в формуле бинома Ньютона для возведения в квадрат, куб, четвертую степень и т.д.

Построенная таблица известна в математике давно и называется *треугольником Паскаля*. Легко заметить, что способ образования строк треугольника Паскаля таков: каждое число равно сумме двух чисел, стоящих над ним. Можно заметить и другие свойства биномиальных коэффициентов исходя из треугольника Паскаля. Приведем некоторые из них (без общих доказательств).

Свойство 1

Биномиальные коэффициенты, равноотстоящие от концов одной строки, равны между собой: $C_n^k = C_n^{n-k}$.

Свойство 2

Сумма биномиальных коэффициентов, стоящих в одной строке, равна степени двойки:

$$\begin{aligned} 1 &= 1 \\ 1+1 &= 2 \\ 1+2+1 &= 2^2 \\ 1+3+3+1 &= 2^3 \\ \ldots\ldots\ldots\ldots \end{aligned}$$

В общем виде

$$C_n^0 + C_n^1 + \ldots + C_n^{n-1} + C_n^n = 2^n.$$

Свойство 3

Биномиальные коэффициенты возрастают от первого до середины, а затем убывают до последнего.

Свойство 4

$$C_{n-1}^{k-1} + C_{n-1}^k = C_n^k.$$

Это свойство записывает в общем виде закон образования новых строк треугольника Паскаля.

Контрольные вопросы и задания

1. Что такое сочетание?
2. Какова формула для числа сочетаний?
3. Как строится треугольник Паскаля?
4. Перечислите свойства биномиальных коэффициентов.

§ 40. Вероятность

Классическое определение вероятности

Выражения «маловероятно», «очень вероятно», «невероятно» часто употребляются в обыденной жизни. Подходя, например, к остановке автобуса, где останавливаются автобусы двух маршрутов — № 2 и 5, и увидев, что автобус № 2 отошел от остановки, а еще один автобус подходит к остановке, вы, конечно, посчитаете более вероятным, что приближающийся автобус имеет номер 5. В более сложных ситуациях нас не может удовлетворить оценка — «одно событие более вероятно, чем другое», необходимо количественно оценить вероятность того или иного возможного события. Необходимость количественной оценки вероятных событий появилась давно и послужила толчком к созданию в XVII—XVIII вв. новой математической дисциплины — теории вероятностей.

Начавшись с оценки вероятности выигрыша или проигрыша в азартных играх, теория вероятностей постепенно пришла к решению трудных задач измерения надежности сложных устройств, эффективности работы различных систем, оценки качества больших партий изделий, задач долгосрочного планирования и многих других прикладных задач.

Простейшей ситуацией, в которой мы без колебаний приписываем событиям определенные числа, является подбрасывание вверх монеты. Здесь мы считаем, что вероятности выпадения герба или цифры (орла или решки) равны по 0,5. (Вместо дроби иногда используют проценты и говорят, что герб выпадает с вероятностью 50%, но мы обычно для исчисления вероятностей будем использовать дроби.)

Ситуация с подбрасыванием монеты действительно очень проста — есть два возможных исхода (выпадение герба или цифры), которые считаются равновероятными, и для определения вероятности каждого из возможных исходов разбиваем единицу (100%) на две равные части. Нетрудно эту ситуацию обобщить на тот случай, когда результаты испытаний можно разбить на некоторое число вариантов (исходов), которые считаются одинаково возможными или равновероятными.

Например, при бросании кубика с нанесенными на его грани числами от 1 до 6 считается равновероятным выпадение любого из этих чисел и каждому такому исходу приписывают вероятность $\frac{1}{6}$. Если вычисляют вероятность более сложного события, которое можно разбить на простейшие равновероятные события, то нужно составить отношение числа «хороших», т.е. благоприятных, исходов к числу всех возможных исходов.

Пример

1. Пусть бросается игральный кубик, и, как уже отмечено, будем считать вероятностью выпадения каждого из чисел от 1 до 6 число $\frac{1}{6}$. Как подсчитать вероятность выпадения числа, делящегося на 3?

   Из всех возможных шести исходов благоприятными являются два: выпадение тройки и выпадение шестерки. Искомой вероятностью будет дробь $\frac{2}{6}=\frac{1}{3}$.
2. Монета бросается три раза подряд. Какова вероятность того, что два раза выпадет герб и, следовательно, один раз цифра?

   Обозначим исход одного бросания Г или Ц (герб или цифра). Возможные исходы трех бросаний получаются обычным перебором: ГГГ, ГГЦ, ГЦГ, ГЦЦ, ЦГГ, ЦГЦ, ЦЦГ, ЦЦЦ. Все эти восемь исходов равновероятны. Из них благоприятными по условию являются ГГЦ, ГЦГ, ЦГГ. Искомая вероятность равна $\frac{3}{8}$.

Вероятностью события называют отношение числа благоприятных исходов к числу всех исходов.

Такое определение вероятности появилось самым первым, по-видимому, еще в XVI в., его обычно называют *классическим определением вероятности*. Оно предполагает, что рассматриваемое испытание может иметь своим результатом некоторое (конечное) число равновероятных исходов.

Применение комбинаторики к решению простейших вероятностных задач

При вычислении вероятности события с помощью классического определения решаются две комбинаторные задачи: определение общего числа равновозможных исходов и определение числа благоприятных исходов.

Задачи

1. С какой вероятностью можно угадать три номера в тираже «Спортлото 5 из 36»?

   Соответствующие комбинаторные задачи мы уже решили раньше. Общее число возможных результатов равно $C_{36}^5 = 376\ 992$. Число благоприятных результатов равно $10 \cdot C_{31}^2 = 4650$. Искомая вероятность равна $\frac{4650}{376\,992} = \frac{775}{62\,832}$.

   Десятичное приближение этой дроби с точностью до 0,001 равно 0,012.

2. В группе 20 человек, 5 из них должны быть выбраны дежурными. С какой вероятностью одному учащемуся не выпадет быть дежурным?

   Общее число вариантов выбора 5 человек из 20 равно C_{20}^5. Для получения благоприятных вариантов надо выбирать 5 человек из 19 (из всех, кроме одного). Число таких вариантов равно C_{19}^5. Для нахождения вероятности не будем отдельно вычислять коэффициенты C_{20}^5 и C_{19}^5, а составим дробь и произведем сокращения. Обозначив искомую вероятность буквой p, получим

$$p = \frac{C_{19}^5}{C_{20}^5} = \frac{19 \cdot 18 \cdot 17 \cdot 16 \cdot 15}{5!} \cdot \frac{5!}{20 \cdot 19 \cdot 18 \cdot 17 \cdot 16} = \frac{15}{20} = \frac{3}{4} = 0{,}75 .$$

   Подумайте, нельзя ли этот же ответ получить более простым рассуждением.

3. Сохраним условие предыдущей задачи и найдем вероятность того, что вам выпадет быть дежурным. Можно, конечно, снова подсчитать число «благоприятных» исходов (проверьте, что оно равно C_{20}^5, и затем найти вероятность по определению (проверьте, что отношение $C_{19}^4 : C_{20}^5$ равно 0,25). Однако ясно, что если вероятность не быть дежурным равна 0,75, то вероятность быть дежурным равна 0,25.

Приведенный пример наводит на мысль, что у вероятностей есть свойства, которые позволяют производить вычисления, не обращаясь каждый раз к комбинаторному классическому определению вероятности. Это действительно так. Попробуйте самостоятельно сформулировать свойство вероятностей, использованное при решении задачи 3.

Повторные испытания

В примере 2 предыдущего подраздела монета бросается три раза подряд. Ситуация, в которой подряд независимо друг от друга производятся одинаковые испытания, встречается очень часто, например бросание монеты или игральной кости, стрельба из одного орудия без учета результата произведенных выстрелов, параллельное включение в сеть одинаковых предохранителей и др.

Разберем более подробно пример с бросанием монеты. При каждом испытании есть два равновероятных исхода: Г (выпал герб) и Ц (выпала цифра). Допустим, что монету бросили подряд n раз. Сколько последовательностей исходов при этом можно получить? Эта комбинаторная задача фактически уже решена была раньше. Последовательность результатов испытаний можно записать как слово в двухбуквенном алфавите, например ГЦЦГЦ. Число n-буквенных слов в таком алфавите равно 2^n. Таким образом, общее число возможных вариантов при повторном бросании монеты будет иметь вид 2^n.

Какова вероятность того, что при n-кратном бросании монеты все время будет выпадать герб? Ясно, что из всех возможных 2^n вариантов благоприятным является один: $\underbrace{\text{ГГ...Г}}_{n \text{ раз}}$. Искомая вероятность равна $\frac{1}{2^n}$. Например, при $n = 10$ эта вероятность равна $\frac{1}{1024} < 0{,}001$.

Интуитивно ясно, что при повторных испытаниях будет получаться результат, стремящийся к нулю. Формула $\frac{1}{2^n}$ дает точную оценку (в опыте с бросанием монеты) интересующей нас вероятности.

Вероятность того, что при бросаниях монеты ни разу не выпадает герб, т.е. все время будет выпадать цифра, тоже равна $1/2^n$.

Возьмем $n = 3$. Ранее простым перебором вариантов вероятность того, что герб выпадает два раза, мы нашли. Если нас интересует только то, сколько раз выпадает герб (или цифра), то возможные варианты будут такими: 1) три раза герб; 2) два раза герб; 3) один раз герб;

4) ни одного раза не выпал герб. Вероятности этих событий равны соответственно $\frac{1}{8}, \frac{3}{8}, \frac{3}{8}, \frac{1}{8}$. Проводя перебор вариантов для $n = 4$, мы получим такой набор вероятностей: $\frac{1}{16}, \frac{4}{16}, \frac{6}{16}, \frac{4}{16}, \frac{1}{16}$. Нетрудно заметить, что в числителях стоят биномиальные коэффициенты. Составим таблицу:

$$
\begin{array}{ccccccccc}
 & & & & 1 & & & & \\
 & & & \frac{1}{2} & & \frac{1}{2} & & & \\
 & & \frac{1}{4} & & \frac{2}{4} & & \frac{1}{4} & & \\
 & \frac{1}{8} & & \frac{3}{8} & & \frac{3}{8} & & \frac{1}{8} & \\
\frac{1}{16} & & \frac{4}{16} & & \frac{6}{16} & & \frac{4}{16} & & \frac{1}{16}
\end{array}
$$

Она получается из треугольника Паскаля делением каждой строки на соответствующую степень двойки, т.е. сумму чисел в этой строке, и представляет таблицу распределения вероятностей различных исходов при повторном бросании монеты. Решите самостоятельно следующие задачи.

1. Какова вероятность того, что при четырех бросаниях монеты хотя бы один раз выпадет герб?
2. Какова вероятность того, что при четырех бросаниях монеты герб выпадет не менее двух раз?
3. Какова вероятность того, что при n бросаниях монеты хотя бы один раз выпадает герб? Запишите последовательность этих чисел в виде десятичных дробей с тремя десятичными знаками.
4. Сколько раз надо бросить монету, чтобы вероятность выпадения хотя бы одного герба была не меньше 0,999?

Выборочный контроль качества продукции

Изделия, выпускаемые цехом, проходят выборочный контроль на прочность: каждое выбранное для проверки изделие ставят под нагрузку. Нагрузку постепенно увеличивают, и изделие признается дефектным, если оно разрушается под нагрузкой 2 т или меньше. Для проверки каждый раз выбирают 25 изделий. Если из них не менее четырех оказались дефектными, то производство останавливается, в противном случае — продолжается.

Обозначим буквой p долю дефектных изделий во всей продукции цеха. Заказчик считает значения $p \le 0{,}05$ допустимыми, а значения $p > 0{,}05$ — недопустимыми. Соответствует ли указанная система контроля требованиям заказчика?

Пусть x — количество дефектных изделий среди выбранных для проверки. Производство останавливается при $x > 3$ и продолжается, если $x \le 3$. Нетрудно подсчитать вероятности этих событий.

Задача

Покажите, что, по аналогии с подбрасыванием монеты, $P\{x = 0\} = (1 - p)^{25}$, $P\{x = 1\} = (1 - p)^{24} p\, C_{25}^{1}$, $P\{x = 2\} = (1 - p)^{23} p^2 C_{25}^{2}$, $P\{x = 3\} = (1 - p)^{22} p^3 C_{25}^{3}$ и поэтому $P\{x \le 3\} = P\{x = 0\} + P\{x = 1\} + P\{x = 2\} + P\{x = 3\}$.

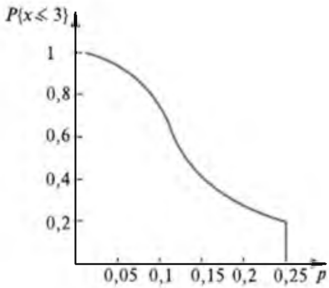


Рис. 385

На рисунке 385 представлена зависимость величины $P\{x \le 3\}$ от p. Ордината точки с абсциссой p дает вероятность того, что производство не будет остановлено. Такой график называется *оперативной характеристикой проверки*.

Если $p < 0{,}05$, то, как показывает график, указанный метод контроля редко приводит к остановке производства. С другой стороны, если вероятность p станет велика по сравнению с 0,05, например $p = 0{,}15$, то более чем в половине случаев будет $x \ge 3$ и система контроля остановит производство.

Однако примерно в половине всех случаев увеличение p от 0,05 до 0,15 может быть не сразу обнаружено. Главный инженер может изменить оперативную характеристику так, чтобы с помощью выборочного контроля быстрее и точнее определить выход p за допустимый уровень 0,05. Для этого достаточно увеличить размер проверяемой партии изделий n (было $n = 25$) и изменить контрольное число r (было $r = 4$). Разумеется, с ростом n увеличивается стоимость проверки.

На рисунке 386 показаны оперативные характеристики для нескольких значений параметров n, r. Во всех случаях производство останавливается при $p = 0{,}07$ с вероятностью около 0,5. При возрастании

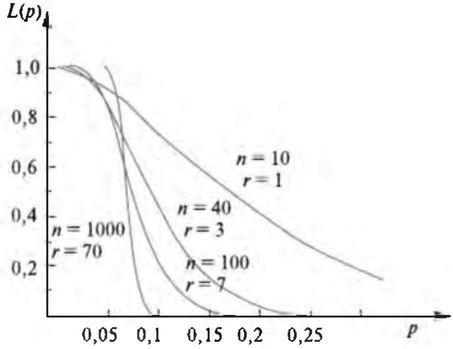


Рис. 386

n величина r тоже возрастает, а самый крутой склон графиков наблюдается около $p = 0{,}07$.

Пользуясь рисунком 385, найдите вероятность того, что производство не будет остановлено при $p = 0{,}1$, при $p = 0{,}3$. Какова вероятность того, что производство будет остановлено, если $p = 0{,}2$? Найдите по рисунку 386 значение p, при котором производство не останавливается в 80% случаев; в 10% случаев.

Постройте оперативную характеристику для $n = 2$, $r = 0$ и сравните ее с рис. 386.

При указанных n, r $P\{x = 0\} = (1 - p)^2$.

Геометрические вероятности

Пусть некоторая плоская фигура разбита на несколько непересекающихся частей. Рассмотрим случайное событие — выбор точки в этой фигуре. Возможными исходами будем считать попадание точки в одну из этих частей. Каждому такому исходу можно приписать вероятность, равную отношению площади части, в которой находится точка, к площади всей фигуры.

Аналогичные определения можно дать для вероятностей попадания точки в заданную часть данного отрезка прямой (или кривой линии), определенную с помощью измерения длины или для случая пространственной фигуры — с помощью объема.

Примеры

1. В квадратном трехчлене $x^2 + 6x + a$ коэффициент a по модулю не больше 10. Он выбирается наудачу. Какова вероятность того, что трехчлен будет иметь вещественные корни?

   Дискриминант d этого трехчлена равен $3^2 - a = 9 - a$. Условие вещественности корней: $d \ge 0 \Leftrightarrow a \le 9$. На отрезке $[-10; 10]$, длина которого равна 20, «благоприятные» значения a занимают отрезок $[-10; 9]$, длина которого 19. Интересующему нас событию следует приписать вероятность $\dfrac{19}{20} = 0{,}95$.

2. В круге произвольно выбирается точка. Какова вероятность того, что ее расстояние до центра круга больше половины радиуса?

Построим две концентрические окружности радиуса R и $\frac{R}{2}$. Площадь маленького круга равна $\frac{1}{4}$ площади большого, а площадь кольца между ними — $\frac{3}{4}$ площади большого. Вероятность попадания точки в кольцо следует принять равной $\frac{3}{4}$.

Заметим, что мы игнорируем границы наших областей. Если вероятность мы измеряем площадью, то вероятность попадания точки на границу области равна нулю, так как площадь границы равна нулю.

3. Часто в приложениях для определения вероятности события приходится строить геометрическую модель и изображать событие точками геометрической фигуры.

Палку ломают случайным образом в двух точках. Какова вероятность того, что из трех получившихся кусков можно составить треугольник?

Построение модели. Пусть длина палки равна 1. Можно сказать, что на единичном отрезке выбирают две точки x и y. Случайное событие выбора пары точек можно изобразить точкой в единичном квадрате на плоскости xOy. Условие того, что из получившихся отрезков можно сложить треугольник, можно записать в виде серии неравенств (будем сразу считать, что x и y обозначены так, что $x < y$) (рис. 387).

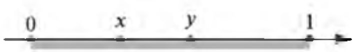


Рис. 387

Эти условия таковы:

1) $x + y - x > 1 - y$;
2) $x + 1 - y > y - x$;
3) $y - x + 1 - y > x$.

Построим область в единичном квадрате, для точек которой выполняются все написанные условия (рис. 388). Мы получим треугольник, площадь которого, как нетрудно проверить, равна $\frac{1}{8}$. Искомая вероятность равна отношению $\frac{1}{8} : \frac{1}{2} = \frac{1}{4}$.

4. Задача Бюффона. Плоскость расчерчена параллельными прямыми, расстояние между которыми равно d. На плоскость случайным образом бросается тонкая игла длины l. Какова вероятность того, что она пересечет какую-либо линию (рис. 389)?

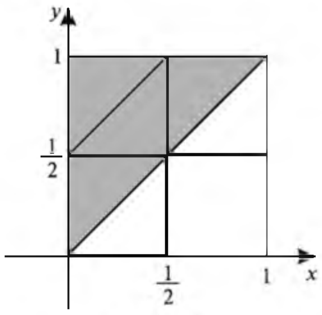


Рис. 388

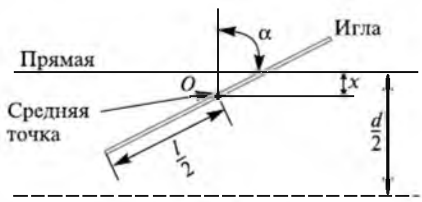


Рис. 389

Опишем случайное событие (бросание иглы) следующими двумя параметрами — положением середины иглы и углом, который игла образует с направлением, перпендикулярным проведенным прямым. Возьмем центр иглы за начало координат O, направим ось Ox перпендикулярно прямым. Из-за симметрии можно считать, что ближайшая к началу прямая расположена над точкой O. Пусть у нас $d = 2$, $l = 1$ Обозначим переменное (случайное) расстояние от точки O до этой прямой через x: $0 \le x \le 1$. Пусть α — угол, образованный иглой с положительным направлением оси Ox (считаем, что $-\frac{\pi}{2} \le \alpha \le \frac{\pi}{2}$). Условие того, что игла пересекает прямую, можно записать в виде неравенства $\frac{1}{2} \cdot \cos\alpha \ge x$. Теперь на плоскости с координатами $(\alpha;\ x)$ изобразим рассматриваемые пары параметров $(\alpha;\ x)$ (рис. 390).

Все возможные пары занимают прямоугольник $-\frac{\pi}{2} \le \alpha \le \frac{\pi}{2}$, $0 \le x \le 1$ площади π. «Благоприятные» пары занимают его часть под графиком функции $x = \frac{1}{2}\cos\alpha$. Площадь S этой части легко находится интегри-

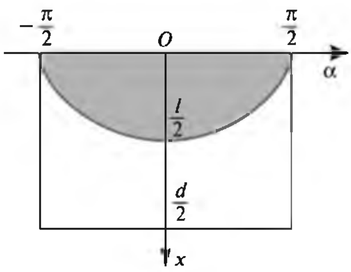


Рис. 390

рованием: $S=\frac{1}{2}\int\limits_{-\frac{\pi}{2}}^{\frac{\pi}{2}}\cos x dx=\frac{1}{2}\sin x\Big|_{-\frac{\pi}{2}}^{\frac{\pi}{2}}=1$. Отношение площадей $\frac{1}{\pi}\approx 0{,}318$ и дает искомую вероятность.

Этот результат пытались многократно проверить экспериментально, бросая иглу и сравнивая частоту, с которой игла пересекает линии, с известным значением числа $\frac{2}{\pi}$.

5. В круге радиуса 1 случайным образом выбирается хорда. Какова вероятность, что ее длина $l \geq 1$?

Возможны различные модели для реализации этого случайного события:

1) фиксируем направление, в котором проводится хорда (в силу симметрии все направления равноправны), проводим радиус, перпендикулярный этому направлению (он пересекает хорду в ее центре), выбираем на радиусе точку — середину хорды. Все точки занимают отрезок длины 1 (радиус круга), «благоприятные» точки — это те, которые отстоят от центра на расстояние, меньшее, чем апофема правильного вписанного шестиугольника, т.е. меньше, чем на $\frac{\sqrt{3}}{2}$ (рис. 391). Искомая вероятность равна $\frac{\sqrt{3}}{2}\approx 0{,}866$;
2) фиксируем один конец хорды и выбираем произвольно второй. Все точки занимают всю окружность, а «благоприятные» точки — две трети этой окружности (рис. 392). Искомая вероятность получилась равной $\frac{2}{3}\approx 0{,}667$;

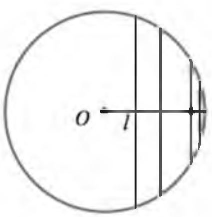


Рис. 391

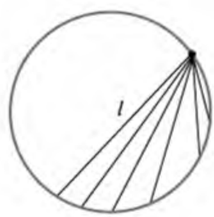


Рис. 392

3) фиксируем направление, в котором проводится хорда. Проведем в этом направлении диаметр и хорду длины 1. Хорда разбила полукруг на две части (рис. 393). Хорды длины, боль-

шей единицы, занимают заштрихованную (нижнюю) часть полукруга. Отношение p площади этой части к площади полукруга легко подсчитать:

$$p = \frac{2}{3} + \frac{\sqrt{3}}{2\pi} \approx 0{,}94 .$$

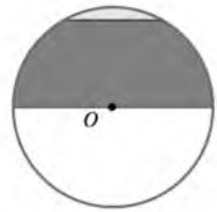


Рис. 393

Мы получили парадоксальную ситуацию. Ее так и называют — парадокс Бертрана (вычисляя вероятность одного и того же случайного события разными способами, мы получили разные ответы). Это показывает, насколько точно должно быть описано случайное событие, и с какой осторожностью надо измерять ее вероятность. В приведенном нами примере *разные* построения хорды описывают *разные* события. Неудивительно, что они имеют разную вероятность.

Математическое ожидание

Представим себе, что в беспроигрышной лотерее разыгрывается шесть призов стоимостью $a_1, a_2, \ldots, a_6$ рублей. Выпадение любого из шести номеров разыгрываемых призов равновероятно. На какой выигрыш в среднем может рассчитывать человек, многократно участвующий в лотерее и каждый раз обладающий одним билетом? Ясно, что ожидаемым выигрышем надо считать среднее арифметическое стоимостей призов, т.е. число $M = \frac{a_1 + \ldots + a_6}{6}$. Это число и называют *математическим ожиданием* выигрыша.

Рассмотрим более сложную ситуацию, когда возможные исходы случайного события не являются равновероятными. Пусть, например, при розыгрыше трех призов стоимостью a_1, a_2 и a_3 вероятности появления их равны соответственно $\frac{1}{2}$, $\frac{1}{6}$ и $\frac{1}{3}$. (То, что $\frac{1}{2} + \frac{1}{6} + \frac{1}{3} = 1$ говорит о том, что должна наступить ровно одна из предполагаемых возможностей). Тогда для определения математического ожидания выигрыша надо числа a_1, a_2 и a_3 взять с весами, равными вероятностям выбора соответствующего приза, т.е. взять число $M = \frac{1}{2}a_1 + \frac{1}{6}a_2 + \frac{1}{3}a_3$.

Вспомним задачу де Мере о справедливом разделе банка. Пусть в игре, состоящей из последовательной серии партий, в каждой из которых выигрывает один из игроков, участвовали три игрока. Выигрывает банк тот, кто первым оказывается победителем в четырех партиях. Игра была прервана после шести партий, в одной из которых выиграл первый игрок, в двух — второй и в трех — третий. Как им справедливо разделить банк? Это и есть задача определения математического ожидания выигрыша для каждого игрока. Ситуация здесь простая — если мы найдем вероятности первым выиграть четыре партии (после уже состоявшихся шести партий) для каждого из игроков, то весь банк и надо разделить пропорционально этим вероятностям.

Итак, первому игроку осталось выиграть три партии, второму две и третьему одну. Попробуем решить задачу просто перечислением вариантов. Будем выписывать в виде «дерева» все равновероятные возможности. Цифра означает номер игрока, выигравшего в очередной партии (рис. 394).

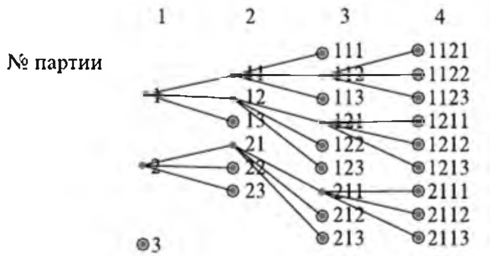


Рис. 394

Кружками обозначены ситуации, когда кто-либо (он указан последним) выиграл четыре партии. Вероятности ситуаций в первой колонке равны $\frac{1}{3}$, во второй — $\frac{1}{9}$, в третьей — $\frac{1}{27}$, в четвертой — $\frac{1}{81}$.

Вероятность выигрыша каждого из участников складывается из вероятностей всех ситуаций, где он завершает игру.

1: $\frac{1}{27}+\frac{3}{81}=\frac{2}{27}$ (он завершает игру в одном случае после трех партий и в трех случаях после четырех).

2: $\frac{1}{9}+\frac{2}{27}+\frac{3}{81}=\frac{2}{9}=\frac{6}{27}$.

3: $\frac{1}{3}+\frac{2}{9}+\frac{3}{27}+\frac{3}{81}=\frac{19}{27}$.

В итоге банк надо разделить в пропорции 2 : 6 : 19. Математические ожидания выигрышей каждым игроком равны $\frac{2A}{27}$, $\frac{6A}{27}$ и $\frac{19A}{27}$, где A — размер банка.

Пусть в результате эксперимента A появляется одно из числовых значений $A_1, A_2, \ldots, A_n$ с вероятностями $p_1, p_2, \ldots, p_n$, $p_1 + p_2 + \ldots + p_n = 1$. Это означает, что определена случайная величина A.

Математическим ожиданием случайной величины A называют число $M = p_1A_1 + p_2A_2 + \ldots + p_nA_n$.

В случае, когда исходы равновероятны, т.е. когда все p_i равны $\frac{1}{n}$, математическое ожидание M равно среднему арифметическому $\frac{A_1 + \ldots + A_n}{n}$.

Контрольные вопросы и задания

1. Что называется вероятностью события?
2. Приведите примеры событий, имеющих вероятность 1, $\frac{1}{2}$, $\frac{1}{3}$, $\frac{1}{6}$, 0.
3. В чем состоит схема повторных испытаний? Приведите примеры.
4. Что такое выборочный контроль качества продукции?
5. Приведите примеры случайных событий, число исходов которых бесконечно.
6. Что такое геометрическая вероятность?
7. Может ли геометрическая вероятность быть иррациональным числом, например, равным $\frac{\sqrt{2}}{2}$?
8. Нарисуйте геометрическую модель для задачи Бюффона.
9. В чем состоит парадокс Бертрана?

ГЛАВА 10

УРАВНЕНИЯ И НЕРАВЕНСТВА

§ 41. Равносильность

Выражение

Уравнение — это самая простая и самая распространенная форма математической задачи.

Выражение — это числа и буквы, соединенные знаками разнообразных операций.

Простейшие арифметические операции (сложение, вычитание, умножение, деление) позволяют составлять выражения такого, например, типа:

$$\frac{2a+b}{a-2b}, \frac{x^2+xy+y^2}{x^2-3xy+y^2} \text{ и др.}$$

Появление новых операций (возведение в степень, логарифмирование, вычисление синуса, тангенса и др.) расширило возможности в составлении выражений. Теперь можно составить более сложные выражения, например такие:

$$\frac{a^{\sqrt{b}}+b^{\sqrt{a}}}{\log_2(a^{1/2}-b^{1/2})}, \frac{\sin x+\operatorname{tg}2y}{\arcsin xy} \text{ и др.}$$

Числа и буквы, входящие в состав выражения, имеют разный смысл. Число, как бы оно ни было записано, например $\frac{1}{2}$; 0,5; 0,4999... или как-то иначе, всегда конкретно, постоянно. Буква же обозначает переменную, меняющуюся величину, которая может принимать разнообразные значения. Мы будем подставлять в выражения вместо букв только числовые значения. При подстановке в выражение вместо букв каких-то чисел мы будем получать *числовые выражения*. Так, числовое выражение $\dfrac{3^2+3\cdot5+5^2}{3^2-3\cdot3\cdot5+5^2}$ получено из выражения $\dfrac{x^2+xy+y^2}{x^2-3xy+y^2}$ подстановкой в него значений $x = 3$, $y = 5$.

Подставляя в выражение определенные значения букв, можно получить числовые выражения, не имеющие смысла. Бессмысленные числовые выражения получаются прежде всего тогда, когда это выражение содержит невыполнимые операции над числами, например деление на нуль, логарифмирование отрицательного числа, арксинус числа, большего единицы, тангенс числа $\dfrac{\pi}{2}$ и др. Другой причиной, приводящей к не имеющим смысла числовым выражениям, является подстановка вместо букв чисел, не входящих в область допустимых значений для этих букв. Например, если в выражении для производительности труда участвует буква a, обозначающая число землекопов в бригаде, то, подставляя значение $a = 2\dfrac{2}{3}$ («два землекопа и две трети»), мы получим бессмысленное числовое выражение, хотя все операции над входящими в выражение числами формально осуществимы.

Областью допустимых значений (ОДЗ) выражения называют множество всех наборов значений букв, при подстановке которых выражение имеет смысл, т.е. превращается в осмысленное числовое выражение.

Заметим, что

если выражение содержит одну букву, то его ОДЗ — это числовое множество, т.е. какое-то подмножество точек числовой прямой. Если же букв, например, две, то ОДЗ выражения — это множество пар чисел и его можно изобразить в виде области, расположенной на координатной плоскости.

Возьмем какое-либо осмысленное числовое выражение и проделаем все указанные в выражении операции над входящими в него числами. Получим одно число — значение числового выражения. Возьмем буквенное выражение и подставим в него вместо букв числа из ОДЗ (т.е. такие числа, чтобы выражение превратилось в осмысленное числовое выражение). Вычислим значение получившегося числового выражения. Это число называют *значением выражения при выбранных значениях букв*. Возможность однозначно вычислить значение выражения при любых допустимых наборах значений входящих в него букв позволяет определить функцию. Вот почему говорят, что выражение можно рассматривать как способ вычисления значений некоторой функции. Поэтому понятие выражения и понятие функции близки между собой.

Два выражения считаются тождественно равными, если равны их числовые значения при любых допустимых наборах значений букв, входящих в эти выражения. Тождество — это два тождественно равных выражения, соединенных знаком равенства.

Примеры тождеств

1. $(a+b)^2 = a^2 \pm 2ab + b^2$.
2. $(a \pm b)^3 = a^3 \pm 3a^2b + 3ab^2 \pm b^2$.
3. $a^2 - b^2 = (a-b)(a+b)$.
4. $a^3 \pm b^3 = (a \pm b)(a^2 \pm ab + b^2)$.

Во всех приведенных тождествах ОДЗ выражений, стоящих слева и справа, совпадают. Часто используют тождества, соединяющие выражения, имеющие разные ОДЗ. В этом случае имеется в виду, что тождество выполняется на общей части ОДЗ выражений, стоящих справа и слева. Поэтому без дополнительных оговорок считаются тождествами следующие равенства.

Примеры тождеств (*продолжение*)

5. $\dfrac{a^2-b^2}{a-b} = a+b$.
6. $(\sqrt{a})^2 = a$.
7. $a^{\log_a x} = x$.
8. $\sin \arcsin x = x$.
9. $\operatorname{tg} x \cdot \operatorname{ctg} x = 1$.
10. $\log_a xy = \log_a x + \log_a y$.

Иногда искусственно (какими-либо дополнительными условиями) уменьшается ОДЗ выражений, составляющих некоторое равенство. Тогда можно говорить о тождестве, выполняющемся на некотором множестве. Так, если $[x]$ обозначает целую часть числа x, то равенство $\left[x+\dfrac{1}{2}\right] = [x]$ является тождеством на множестве целых чисел (но, разумеется, не является тождеством в обычном смысле слова).

Примеры тождеств (*продолжение*)

11. $\arcsin(\sin x) = x$ — тождество на промежутке $\left[-\dfrac{\pi}{2}; \dfrac{\pi}{2}\right]$.
12. $\sqrt{x^2} = x$ — тождество на промежутке $[0; +\infty]$.
13. $\arcsin(\sin x) = \pi - x$ — тождество на промежутке $\left[-\dfrac{\pi}{2}; \dfrac{3\pi}{2}\right]$.

Тождественное преобразование выражения — это переход от одного выражения к тождественно равному выражению.

Самые «безобидные» тождественные преобразования, например приведение подобных членов, сокращение дробей, использование свойств степени и т.п., могут привести к выражению, у которого ОДЗ больше или меньше, чем у исходного выражения. Это может оказаться существенным при решении уравнений, поэтому информацию об изменении ОДЗ при тождественных преобразованиях полезно хранить в памяти (собственной, машинной или просто в тетради).

Уравнение

Возьмем два числовых выражения и поставим между ними знак равенства. Мы получим *числовое равенство*. Оно будет верным или неверным в зависимости от того, равны или не равны значения взятых числовых выражений. Классическими примерами являются равенства $2 \cdot 2 = 4$ и $2 \cdot 2 = 5$.

Отметим еще раз, что когда мы говорим «равенство двух числовых выражений», мы вовсе не утверждаем, что эти два выражения действительно равны. Соединить два числовых выражения A и B знаком «=» и говорить о получившемся равенстве $A = B$ можно независимо от того, верно или неверно сформулированное нами утверждение «$A = B$».

Возьмем два буквенных выражения и соединим их знаком равенства. Получим уравнение. Таким образом,

уравнение в первом приближении можно понимать как равенство двух буквенных выражений.

Равенство числовых выражений иногда называют *безусловным равенством*, т.е. равенством или безусловно верным, или безусловно неверным. Уравнение с этой точки зрения можно считать «*условным равенством*» — при одних условиях (т.е. при одних значениях букв) оно может оказаться верным, при других — неверным. Тождество — это равенство, верное при всех допустимых значениях букв. Его тоже можно считать частным случаем уравнения.

Уравнение — это не просто формальное равенство двух выражений. Главное в понятии уравнения — это постановка вопроса о его решении. Следовательно, уравнение — это равенство двух выражений вместе с призывом найти его решения. Опишем более точно, что же значит решить уравнение.

Буквы, входящие в состав уравнения (т.е. в состав выражений, образующих уравнение), называются *неизвестными*. Если такая буква одна, то говорят, что мы имеем дело с уравнением с одним неизвестным. Аналогично можно говорить об уравнении с двумя, тремя и любым другим числом неизвестных.

Рассмотрим уравнение с одним неизвестным.

Значение неизвестного, при подстановке которого уравнение превращается в верное числовое равенство, называется корнем уравнения. Решить уравнения с одним неизвестным — значит найти все его корни.

Возьмем уравнение с числом неизвестных, большим одного. Например, рассмотрим уравнение с двумя неизвестными. Чтобы получить из него числовое равенство, надо каждому неизвестному придать определенное числовое значение, т.е. взять пару чисел. Решить уравнение с двумя неизвестными — значит найти все пары чисел, удовлетворяющих этому уравнению, т.е. такие, при подстановке которых уравнение превращается в верное числовое равенство. Одну такую пару тоже можно было бы назвать корнем уравнения, но обычно так не говорят, а вводят понятие «решение уравнения».

Решение уравнения с двумя неизвестными — это пара чисел, удовлетворяющих этому уравнению.

Разумеется, и в случае уравнения с одним неизвестным можно вместо слов «корень уравнения» говорить «решение уравнения». Путаница может возникнуть из-за разного употребления слова «решение». Можно сказать о решении уравнения как его корне. При таком употреблении этого слова имеют смысл такие фразы, как «уравнение имеет одно решение», «уравнение имеет три решения», «уравнение не имеет решений». В речи часто используют словосочетание «решение уравнения» как процесс нахождения его корней (решений). Можно сказать так: «уравнение имеет сложное решение», «я не смог найти путь решения этого уравнения». В процессе решения уравнения может обнаружиться, что оно совсем не имеет корней (решений). И в этом случае мы решили уравнение: доказали, что у него решений нет.

Что означает найти корни уравнения? В школьной практике при решении уравнений принято записывать ответ как результат знакомых операций над числами, например:

$$x = \sqrt{2},\ x = \sqrt[3]{1+\sqrt{3}},\ x = \frac{\log_3 7}{\log_3 5},\ x = \sin\frac{\pi}{5},\ x = 2^{\sqrt{3}},\ x = \arcsin\frac{1}{3}.$$

В то же время при решении прикладных задач бывает необходимо представить ответ в десятичной записи с определенным числом знаков после запятой. Такой ответ можно получить, используя калькулятор или другое вычислительное устройство.

Мы условились понимать под уравнением равенство, составленное из двух выражений. Мы уже говорили о том, что выражение можно рассматривать как способ задания некоторой функции. Поэтому

уравнение можно понимать как равенство, соединяющее две функции.

Пусть даны две функции от переменной x, например, $y = f(x)$ и $y = g(x)$. Составим уравнение $f(x) = g(x)$. Оно получено приравниванием выражений $f(x)$ и $g(x)$. Пусть $D_1 = D(f)$ и $D_2 = D(g)$ — области определения функций f и g. Тогда D_1 и D_2 можно понимать как области допустимых значений выражений $f(x)$ и $g(x)$. Общая часть областей, т.е. множество $D_3 = D_1 \cap D_2$, является областью допустимых значений уравнения $f(x) = g(x)$.

Полезно помнить, что подставлять в уравнение можно любое значение x. При каком-то значении x может получиться бессмысленное числовое выражение, а при x из ОДЗ получится осмысленное числовое равенство. Если при этом оно окажется еще и верным, то взятое число x является корнем уравнения.

Вернемся к вопросу о решении уравнения. Начнем с уравнения с одним неизвестным x. В какой форме рекомендуется записывать его ответ?

Уравнение может иметь один корень, например $x = 5$. Тогда ответ проще всего записать в форме $x = 5$.

Уравнение может иметь несколько корней. Ответ удобно записать в виде перечисления всех корней, используя «нумерованные» значения x. Например, $x_1 = -1$, $x_2 = 0$, $x_3 = 1$. Полезно корни располагать в порядке возрастания.

Уравнение может и не иметь корней. В таком случае нагляднее всего это и указать в ответе: корней нет.

Тригонометрические уравнения (и вообще уравнения с периодическими функциями) часто имеют бесконечно много корней, которые можно записать в виде одной или нескольких последовательностей, зависящих от натурального аргумента. Возможна такая запись ответа:

$$x = \frac{\pi}{4} + \pi k,\ k \in Z.$$

Встречаются уравнения, решения которых заполняют один или несколько промежутков, которые и указываются в ответе, например $0 \le x \le 1$, или x — любое число.

Все корни (решения) уравнения образуют множество корней. Слово «множество» не означает, что корней очень много («великое множество»). Если множество корней обозначить одной буквой, например X, то ответ может быть записан иначе. Примеры записи ответов с употреблением теоретико-множественных обозначений: $X = \{5\}$; $X = \{-1; 0; 1\}$; $X = \varnothing$ (пустое множество, т.е. корней нет; не надо путать знак пустого множества с обозначением нуля); $X = [0; 1]$; $X = \left\{\frac{\pi}{4} + \pi k \mid k \in Z\right\}$; $X = \mathbb{R}$.

Множество решений уравнения с двумя неизвестными состоит из пар значений этих неизвестных. Важно помнить, что одна пара, например $x = 1$, $y = 5$, — это одно решение, а не два.

Равносильность

Если идет дождь, то я открываю зонт. Можно сказать, что открывание зонта является следствием того, что идет дождь. Если число делится на 6, то оно четно. Так же как и раньше, можно сказать, что четность числа является следствием его делимости на 6.

Пусть даны два уравнения, которые мы назовем уравнение A и уравнение B. Если каждый корень уравнения A является корнем уравнения B, то уравнение B является следствием уравнения A: $A \Rightarrow B$ (читается: «из A следует B», или «B является следствием A», или «если A, то B»).

На языке теории множеств можно сказать короче: уравнение B является следствием уравнения A, если множество корней уравнения A содержится в множестве корней уравнения B, т.е. если $X_A \subset X_B$, где X_A и X_B — упомянутые множества.

Переходя от одного уравнения к его следствию, мы не потеряем корней исходного уравнения, но, возможно, приобретем лишние. Основой получения разнообразных следствий является следующее простое соображение. Пусть $a = b$ — числовое равенство, а f — функция, определенная в точках a и b. Тогда равенство $f(a) = f(b)$ является следствием равенства $a = b$, т.е. если равенство $a = b$ верно, то верно и равенство $f(a) = f(b)$ (если оно имеет смысл).

Возьмем теперь уравнение, полученное приравниванием двух выражений.

Если функция f определена при всех значениях этих выражений, то, применяя функцию f к обеим частям уравнения, получим новое уравнение, являющееся следствием исходного.

Это правило особенно удобно, если функция f определена при любых числовых значениях переменных.

Примеры

Возьмем уравнение $\sqrt{x+1}=x^2+1$.

Следующие уравнения являются его следствиями (рядом записана применяемая функция, а буквой z обозначен ее аргумент).

1. $(\sqrt{x+1})^2=(x^2+1)^2$; $f(z)=z^2$.
2. $\sqrt{x+1}-1=x^2$; $f(z)=z-1$.
3. $2^{\sqrt{x+1}}=2^{x^2+1}$; $f(z)=2^z$.
4. $\sin(\sqrt{x+1})=\sin(x^2+1)$; $f(z)=\sin z$.

   Все функции f определены при любом z, поэтому получение указанных следствий было формальной операцией.
5. $\log_2\sqrt{x+1}=\log_2(x^2+1)$; $f(z)=\log_2 z$.
6. $\sqrt[4]{x+1}=\sqrt{x^2+1}$; $f(z)=\sqrt{z}$.
7. $\dfrac{1}{\sqrt{x+1}}=\dfrac{1}{x^2+1}$; $f(z)=\dfrac{1}{z}$.
8. $\arcsin\dfrac{1}{2}\sqrt{x+1}=\arcsin\dfrac{1}{2}(x^2+1)$; $f(z)=\arcsin\dfrac{z}{2}$.

В случаях 5—8 функции уже определены не при всех x. Однако во всех случаях каждое новое уравнение является следствием исходного. Этот вывод уже не является формальным. Примеры 5—7 разберите самостоятельно. Пример 8 существенно более трудный и требует дополнительных сведений о корнях исходного уравнения (докажите, что все его корни лежат на отрезке $[0; 1]$).

Два уравнения называются равносильными, если каждое из них является следствием другого, т.е. если каждый корень одного из них является корнем другого.

Пусть уравнение A имеет множество корней X_A, а уравнение B — множество X_B. Равносильность уравнений A и B обозначается так: $A \Leftrightarrow B$. По определению, равносильность означает выполнение двух условий: $A \Rightarrow B$ (уравнение B является следствием уравнения A) и $B \Rightarrow A$ (наоборот, уравнение A является следствием уравнения B). На языке теории множеств равносильность означает равенство $X_A = X_B$.

Итак, у равносильных уравнений корни одни и те же. Поэтому основным способом решения уравнения является следующий: с помощью перехода от одного уравнения к равносильному стараются прийти к уравнению, решения которого находятся легко.

Основной способ получения следствия нам известен — применение какой-либо функции к обеим частям уравнения.

Чтобы этот переход сохранял равносильность, надо, чтобы возможен был обратный переход. Это всегда выполняется, если новое уравнение получено с помощью функции, имеющей обратную. На этом основаны теоремы о равносильности, позволяющие утверждать равносильность пар уравнений, получающихся друг из друга с помощью взаимно обратных функций. Сформулируем несколько таких теорем.

Запишем уравнение в символической форме:

$$\square = \triangle,$$

где $\square$ и $\triangle$ — два выражения, составляющие уравнение.

Во всех этих случаях не было трудностей с областями определения применяемых функций. Использование таких распространенных функций, как возведение в квадрат, умножение и деление на некоторую функцию, нахождение обратной величины и других, в общем виде не гарантирует равносильности. Например, возводя в квадрат обе части уравнения, мы получаем следствие:

$$\square = \triangle \Rightarrow \square^2 = \triangle^2.$$

В общем виде обратный переход неверен. Однако если из последующего решения уравнения $\square^2 = \triangle^2$ мы узнаем, что для его корней выражения $\square$ и $\triangle$ имеют одинаковый знак, то можно поставить стрелку в обратном направлении и найти корни исходного уравнения:

$$\square^2 = \triangle^2 \Rightarrow \square = \triangle,$$

если $\square$ и $\triangle$ одного знака. Представим основные случаи, при которых уравнение переходит в эквивалентное (табл. 10).

Таблица 10

Переход уравнений

Случай	Уравнение	Эквивалентное уравнение
1	$\square = \triangle$ $\Updownarrow$ $\square + a = \triangle + a$	$z = y - a$ $\downarrow \quad \uparrow$ $z + a = y$
2	$\square = \triangle$ $\Updownarrow \qquad a \neq 0$ $a \cdot \square = a \cdot \triangle$	$z = \frac{1}{a} y$ $\downarrow \quad \uparrow$ $az = y$

Окончание

Случай	Уравнение	Эквивалентное уравнение
3	$\square = \triangle$ $\Updownarrow$ $\quad a > 0,$ $a^{\square} = a^{\triangle}$ $\quad a \neq 1$	$z = \log_a y$ $\downarrow \ \uparrow$ $a^z = y$
4	$\square = \triangle$ $\Updownarrow$ $\square^3 = \triangle^3$	$z = \sqrt[3]{y}$ $\downarrow \ \uparrow$ $z^3 = y$

Остановимся подробнее на некоторых полезных преобразованиях уравнений.

1. Тождественное преобразование одной из частей уравнения и перенос членов из одной части уравнения в другую с противоположным знаком приводит к равносильному уравнению, если при этом не происходит изменения ОДЗ.

Пример

Уравнение $\dfrac{x}{x^2+2} = \dfrac{1}{3}$ равносильно уравнению $x^2 - 3x + 2 = 0$.

В то же время уравнения

$\lg (x-2)(x+6) = 1 + \lg 2$,

$\lg (x-2) + \lg (x+6) = 1 + \lg 2$

не являются равносильными (корни первого $x_1 = -8$, $x_2 = 4$; корень второго $x = 4$), так как логарифмирование произведения уменьшило ОДЗ.

2. *Совокупность уравнений.* Рассмотрим задачу, в которой требуется решить несколько уравнений, а затем объединить их корни. Можно сказать, что идет речь о решении совокупности уравнений. Обычно совокупность обозначается с помощью прямой скобки.

Пусть ОДЗ выражений $\square$ и $\triangle$ совпадают. Тогда уравнение $\square \cdot \triangle = 0$ равносильно совокупности $\left[\begin{array}{l}\square = 0,\\ \triangle = 0.\end{array}\right.$

Оговорка при совпадении ОДЗ не случайна. Так, уравнение $\cos x \operatorname{tg} x = 0$ не равносильно совокупности $\left[\begin{array}{l}\cos x = 0,\\ \operatorname{tg} x = 0.\end{array}\right.$

3. *Система уравнений.* Рассмотрим задачу, в которой надо решить несколько уравнений и взять их общие корни (или, иначе, найти числа, удовлетворяющие каждому из уравнений систе-

мы). В систему можно объединить не только уравнения, но и различные условия, ограничения неравенства. Например, решить систему

$$\begin{cases} x^2+2x-3=0, \\ x+1>0 \end{cases}$$

означает, что надо решить первое уравнение и взять только те его корни, для которых выполняется неравенство $x+1>0$. Использование переходов от уравнения к совокупностям и системам позволяет разнообразить схемы равносильных переходов. Покажем некоторые из них.

1. $$\frac{\square}{\triangle}=0 \Leftrightarrow \begin{cases} \square=0, \\ \triangle\neq 0; \end{cases}$$

2. $$\frac{1}{\square}=\frac{1}{\triangle} \Leftrightarrow \begin{cases} \square=\triangle, \\ \square\neq 0; \end{cases}$$

3. $$\square=\triangle \Leftrightarrow \left[\begin{array}{l} \begin{cases} \square^2=\triangle^2, \\ \square\geq 0, \\ \triangle\geq 0; \end{cases} \\ \begin{cases} \square^2=\triangle^2, \\ \square<0, \\ \triangle<0; \end{cases} \end{array}\right.$$

4. $$\square=\triangle \Leftrightarrow \left[\begin{array}{l} \begin{cases} \lg\square=\lg\triangle, \\ \square>0; \end{cases} \\ \begin{cases} \lg(-\square)=\lg(-\triangle), \\ \square<0; \end{cases} \\ \begin{cases} \square=\triangle, \\ \triangle=0. \end{cases} \end{array}\right.$$

Неравенство

Почти все, что было выше сказано об уравнении, можно дословно перенести и на неравенство. Прежде всего, отметим, что знаков неравенства четыре: $>$ (больше), $<$ (меньше), $\geq$ (больше или равно), $\leq$ (меньше или равно). Мы будем говорить о каком-либо одном из них.

Числовое неравенство получается соединением двух числовых выражений знаком неравенства. Аналогично равенствам, числовые неравенства могут быть верными и неверными. В приведенных далее примерах все неравенства с нечетными номерами являются верными, а с четными — неверными.

Примеры

1. $3>2$.	3. $3\geq 3$.	5. $-3<5$.	7. $3\leq 5$.
2. $-3>-1$.	4. $2\geq 3$.	6. $2<2$.	8. $-1\leq -2$.

Приведем основные правила преобразования неравенств, используя знак следствия $\Rightarrow$ и равносильности $\Leftrightarrow$.

1. $a > b \Leftrightarrow a - b > 0 \Leftrightarrow b < a \Leftrightarrow b - a < 0$.
2. $a > b \Leftrightarrow a + c > b + c$.
3. $\begin{cases} a > b \\ c > 0 \end{cases} \Rightarrow ac > bc$.
4. $\begin{cases} a > b \\ c < 0 \end{cases} \Rightarrow ac < bc$.
5. $a > b > 0 \Rightarrow \dfrac{1}{a} < \dfrac{1}{b}$.
6. $0 > a > b \Rightarrow \dfrac{1}{a} < \dfrac{1}{b}$.

Основой техники преобразования неравенств является следующее общее соображение:

Пусть функция f монотонна на промежутке, содержащем числа a и b. Тогда $a < b \Rightarrow f(a) < f(b)$, если f строго возрастает: $a < b \Rightarrow f(a) > f(b)$, если f строго убывает.

Свойства 3—6 получаются применением этого правила к функциям $y = cz$ и $y = \dfrac{1}{z}$. Аналогично, для функций $y = z^2$ и $y = 2^z$ можно записать:

7. $a > b > 0 \Rightarrow a^2 > b^2$.
8. $0 > a > b \Rightarrow a^2 < b^2$.
9. $a > b \Rightarrow 2^a > 2^b$.

Неравенство с одним неизвестным получается при соединении знаком неравенства двух выражений, содержащих одну букву или что близко по смыслу, двух функций от одной и той же переменной. Аналогично можно рассматривать неравенства с двумя (и более) неизвестными.

Ограничимся неравенствами с одним неизвестным.

Область допустимых значений неравенства — это множество значений неизвестного, при подстановке которых получается осмысленное числовое неравенство. Решение неравенства — это такое значение неизвестного, при подстановке которого получается верное числовое неравенство. Решить неравенство — это значит найти, описать множество его решений.

Два неравенства называются равносильными, если множества их решений совпадают. Одно неравенство является следствием другого, если множество его решений содержит в себе множество решений второго. Ясно, что каждое из равносильных неравенств является следствием другого.

Технику решения неравенств с помощью переходов, сохраняющих равносильность, мы покажем на примерах.

Параметр

Запишем уравнение $ax^2 + bx + c = 0$. Выражение, стоящее в его левой части, содержит четыре буквы: x, a, b, c. Хотя все эти четыре буквы равноправны, мы смотрим на это уравнение как на квадратное уравнение относительно неизвестного x, считая a, b, c буквенными коэффициентами, ***параметрами***. Необходимость рассматривать уравнение с буквенными коэффициентами возникает часто. Прежде всего, это полезно тогда, когда формулируются некоторые общие свойства, присущие не одному определенному уравнению, а целому классу уравнений. Так можно сформулировать свойства корней квадратного уравнения, показательного уравнения $a^x = b$, тригонометрического уравнения $\sin \omega x = a$ в зависимости от параметров a, b, ω.

Разумеется, то, что в уравнении одни буквы мы считаем неизвестными, а другие — параметрами, в значительной степени условно. В реальной практике из одного и того же соотношения между переменными приходится выражать одни переменные через другие, т.е. решать уравнение относительно одной буквы, считая ее обозначением неизвестного, а другие буквы — параметрами.

По традиции неизвестные обозначаются последними буквами латинского алфавита — x, y, z, а параметры — первыми — a, b, c, или вообще буквами другого алфавита (например, греческими).

При решении уравнений и неравенств с параметрами чаще всего встречаются две задачи.

Задачи

1. Найти формулы для решений уравнения (неравенства), выражающие эти решения как функции от параметров. Типичный пример — формула корней квадратного уравнения.
2. Исследовать решения уравнения (неравенства) в зависимости от изменения значений параметров. Скажем, очень часто встречается такая задача — найти число корней уравнения в зависимости от параметра или определить, при каких значениях параметра уравнение не имеет корней. Очень часто исследование корней в зависимости от параметра можно провести, не вычисляя самих корней.

Пример

Дано уравнение $x^2 + 2x + a = 0$ относительно неизвестного x с параметром a.

1. При каких значениях a уравнение имеет два корня?
2. При каких значениях a уравнение имеет два корня, причем один из них больше единицы, а другой меньше?
3. При каких значениях a сумма квадратов корней меньше шести?

Решите этот пример самостоятельно.

Ответы: 1) $a < 1$; 2) $a < -3$; 3) $-1 < a < 1$.

Контрольные вопросы и задания

1. Проверьте, понимаете ли вы смысл следующих ключевых слов и обозначений: выражение, тождество, уравнение, корень, решение уравнения, неравенство, решение неравенства, $\Rightarrow$,$\Leftrightarrow$, ОДЗ.
2. Что такое ОДЗ выражения?
3. Приведите примеры верных и неверных числовых равенств.
4. Что такое решение уравнения с тремя неизвестными?
5. Приведите пример уравнения, имеющего единственное решение.
6. Приведите пример уравнения, имеющего более одного, но конечное множество решений.
7. Приведите пример уравнения, имеющего бесконечно много решений.
8. Приведите пример уравнения, не имеющего решений.
9. Что означает, что одно уравнение является следствием другого?
10. Одно уравнение имеет два корня: $x_1 = 1, x_2 = 3$. Какие корни может иметь второе уравнение, чтобы первое уравнение было его следствием?
11. Какие уравнения называются равносильными?
12. Первое уравнение является следствием второго, второе — следствием третьего, а третье — следствием первого. Что можно сказать о равносильности этих уравнений?
13. Что может произойти с уравнением, если мы обе его части возведем в квадрат? Приведите примеры.
14. Что может произойти с уравнением, если мы прологарифмируем обе его части? Приведите примеры.
15. Чем отличается совокупность уравнений от системы уравнений?
16. Могут ли быть одновременно верными числовые неравенства $a < b$ и $a > b$?

17. Какие неравенства называются равносильными?
18. Что может произойти при возведении обеих частей неравенства в квадрат?
19. Если некоторое неравенство вида $A < B$ не имеет решения, то что можно сказать о решении неравенства $A \geq B$?
20. Решением какого уравнения можно заменить решение системы неравенств $\begin{cases} A \geq B, \\ B \geq A? \end{cases}$

§ 42. Уравнения с одним неизвестным

Общие приемы

В простейших случаях решение уравнения с одним неизвестным распадается на два шага — преобразование уравнения к стандартному и решение стандартного уравнения. Второй шаг осуществляется по известным формулам, которые всегда можно восстановить в памяти с помощью справочников. Есть они и в данном учебнике — это стандартные уравнения, которые были нами изучены.

1. Линейное уравнение $ax + b = 0$.
2. Квадратное уравнение $ax^2 + bx + c = 0$.
3. Простейшее степенное уравнение $x^n = a$.
4. Показательное уравнение $a^x = b$.
5. Логарифмическое уравнение $\log_a x = b$.
6. Простейшие тригонометрические уравнения $\sin x = a$, $\cos x = a$, $\operatorname{tg} x = a$, $\operatorname{ctg} x = a$.

Преобразование уравнения к одному из стандартных является основным шагом в решении уравнения. Полностью алгоритмизировать процесс преобразования нельзя, однако полезно запомнить некоторые наиболее употребительные приемы, общие для всех типов уравнений.

1. *Разложение на множители*. Если уравнение равносильными преобразованиями удается привести к виду $\square \cdot \triangle = 0$, то исходное уравнение равносильно совокупности двух более простых уравнений $\left[\begin{array}{l} \square = 0, \\ \triangle = 0 \end{array}\right.$ при условии сохранения ОДЗ.

Этот прием часто применяется при решении алгебраических уравнений степени выше второй, при решении тригонометрических уравнений. Соответствующие примеры будут приведены далее.

2. *Введение нового неизвестного.* Посмотрите, не решая, на следующий набор уравнений:

1. $(x^2 + 3x)^2 + 2(x^2 + 3x) - 120 = 0.$
2. $\sqrt{x^2+3x} + \sqrt{x^2+3x+1} = 1.$
3. $2^{x^2+3x} - 2^{x^2+3x-1} = \dfrac{1}{2}.$
4. $\log_2^2(x^2 + 3x) - \log_2(x^2 + 3x) = 2.$

В каждом из этих уравнений отметим присутствие выражения $x^2 + 3x$. Если заменить его буквой y ($y = x^2 + 3x$), то получим более простые уравнения относительно y:

1. $y^2 + 2y - 120 = 0.$
2. $\sqrt{y} + \sqrt{y+1} = 1.$
3. $2^y - 2^{y-1} = \dfrac{1}{2}.$
4. $\log_2^2 y - \log_2 y = 2.$

Найдя из этих уравнений значения y, подставим их в соотношение $y = x^2 + 3x$ и вычислим корни исходного уравнения.

3. *Графический метод.* Рассмотрим уравнение с одним неизвестным: $f(x) = g(x)$. Изобразим на одном рисунке графики функций $y = f(x)$ и $y = g(x)$ (рис. 395). Точкам пересечения графиков этих функций соответствуют те значения аргумента x, при которых совпадают значения функций, т.е. корни данного уравнения.

Итак, абсциссы точек пересечения графиков функций $y = f(x)$ и $y = g(x)$ являются корнями уравнения $f(x) = g(x)$. Например, для уравнения $x^2 = x + 2$ такими точками будут $P_1 = (-1; 1)$ и $P_2(2; 4)$, т.е. $x_1 = -1$, $x_2 = 2$ (рис. 396).

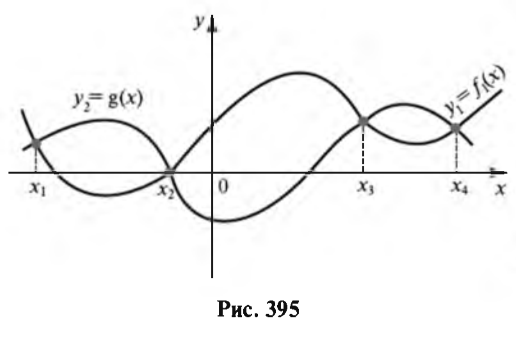


Рис. 395

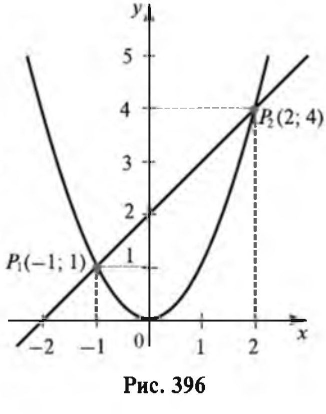


Рис. 396

Если уравнение имеет вид $f(x) = 0$, то в качестве функции, стоящей в правой части, выступает функция $y = 0$. Графиком ее будет ось x, поэтому корнями уравнения $f(x) = 0$ являются абсциссы точек пересечения графика функции $y = f(x)$ с осью x (рис. 397).

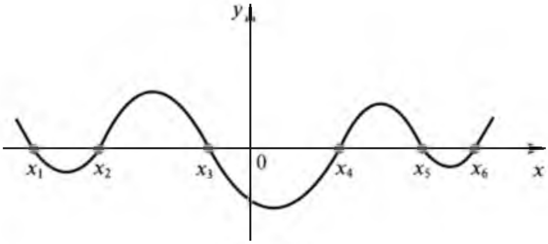


Рис. 397

Графическая иллюстрация решения уравнения указывает, на первый взгляд, и способ решения уравнения: строят в системе координат две кривые и находят их точки пересечения. Действительно, если выбрать масштаб и построить графики достаточно аккуратно, то можно приближенно найти точки пересечения и их абсциссы — корни уравнения. Но для того чтобы найти координаты точек пересечения точно, как раз и нужно решить соответствующее уравнение! В то же время графическая иллюстрация часто дает некоторые качественные ответы, количество корней, а также грубо указывает отрезки на числовой оси, где эти корни могут находиться. Рассмотрим в качестве примера уравнение $(x - 1)^2 = \sqrt{x}$.

Построим графики функций, стоящих в левой и правой частях.

Из рисунка 398 можно заключить, что уравнение имеет два корня, один из которых находится в интервале (0; 1), а другой — в интервале (2; 3). Можно указывать эти интервалы и более точно: (0; 0,5) и (2; 2,5), еще более точно: (0,2; 0,3) и (2,2; 2,3). (Действительно, нетрудно проверить, что при $x = 0{,}2$ имеем $\sqrt{x} < (x - 1)^2$, а при $x = 0{,}3$ — уже $\sqrt{x} > (x - 1)^2$; точно так же при $x = 2{,}2$ левая часть уравнения больше правой, а при $x = 2{,}3$ — меньше). Вообще, вычисляя и сравнивая значения левой и правой частей уравнения, можно найти корни с любой степенью точности.

Корни уравнения пятой степени $x^5 - 3x + 1 = 0$ вообще нельзя записать с помощью радикалов, но, построив достаточно точный график функции $y = x^5 - 3x + 1$ (рис. 399), можно определить, что уравнение имеет три корня в интервалах $(-1{,}5; -1{,}3)$, $(0; 0{,}5)$ и $(1; 1{,}3)$.

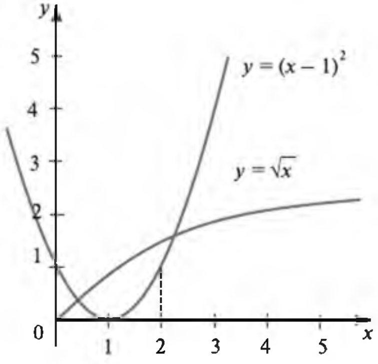


Рис. 398

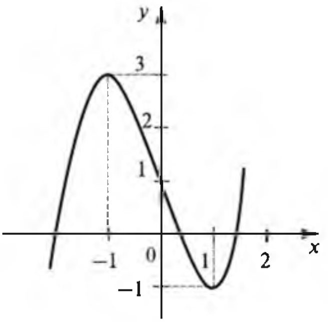


Рис. 399

Примеры решения уравнений

1. *Алгебраическое уравнение* $x(x+1)(x+2)(x+3) = 120$. Если раскрыть скобки и привести подобные члены, то получится уравнение четвертой степени. Общий прием решения уравнения четвертой степени нам неизвестен, поэтому не будем торопиться раскрывать скобки.

   Способ 1. Воспользуемся симметрией левой части. Перемножив первый и четвертый множители, а также второй и третий, получим $(x^2+3x)(x^2+3x+2) = 120$. После замены $x^2+3x = y$ уравнение сводится к квадратному $y(y+2) = 120$.

   Способ 2. Симметрией можно воспользоваться иначе. Заметим, что числа x, $x+1$, $x+2$, $x+3$, расположены на числовой оси симметрично относительно числа $x + \frac{3}{2}$. Сделаем замену $x + \frac{3}{2} = y$. Тогда $x = y - \frac{3}{2}$, $x+1 = y - \frac{1}{2}$, $x+2 = y + \frac{1}{2}$, $x+3 = y + \frac{3}{2}$ и $\left(y-\frac{3}{2}\right)\left(y-\frac{1}{2}\right)\left(y+\frac{1}{2}\right)\left(y+\frac{3}{2}\right)=120$.

   Теперь преобразования более очевидны $\left(y^2-\frac{9}{4}\right)\left(y^2-\frac{1}{4}\right)=120$.

   Это биквадратное уравнение, приводящееся к квадратному заменой.

   Способ 3. Перемножив все скобки, получим уравнение $x^4+6x^3+11x^2+6x-120=0$. Попробуем подобрать корень. Легко догадаться, что $2\cdot3\cdot4\cdot5=120$, поэтому подстановкой проверяем, что $x=2$ является корнем. Выделим множитель $x-2$:

$x^4 + 6x^3 + 11x^2 + 6x - 120 = x^4 - 2x^3 + 8x^3 - 16x^2 + 27x^2 - 54x + 60x - 120 = (x-2)(x^3 + 8x^2 + 27x + 60)$.

Теперь подбираем корень уравнения $x^3 + 8x^2 + 27x + 60 = 0$. Можно угадать $x = -5$ (так как $(-5)(-4)(-3)(-2) = 120$). Выделим множитель $x + 5$:

$x^3 + 5x^2 + 3x^2 + 15x + 12x + 60 = (x+5)(x^2 + 3x + 12)$.

У оставшегося квадратного трехчлена $x^2 + 3x + 12$ вещественных корней нет.

Способ 4. Объединим первый множитель с последним, а второй с третьим:

$x(x+1)(x+2)(x+3) = x(x+3)\ (x+1)(x+2) = (x^2 + 3x) \times (x^2 + 3x + 2)$.

Обозначим

$(x^2 + 3x) = u \Rightarrow (x^2 + 3x)(x^2 + 3x + 2) = u(u+2) = u^2 + 2u + 1 - 1 = (u+1)^2 - 1$.

Таким образом мы получили тождество:

$x(x+1)(x+2)(x+3) = (x^2 + 3x + 1)^2 - 1$.

Отсюда получаем

$$(x^2 + 3x + 1)^2 = 121 \Leftrightarrow \left[\begin{array}{l} x^2 + 3x + 1 = 11, \\ x^2 + 3x + 1 = -11. \end{array}\right.$$

Ответ: $x_1 = -5$, $x_2 = 2$.

2. ***Уравнение с модулем***: $|x^2 + 2x| + x^2 + x = 5$.

   Уравнение равносильно совокупности двух систем:

   $$\begin{cases} x^2 + 2x + x^2 + x = 5 \\ x^2 + 2x \geq 0; \end{cases} \quad \begin{cases} -x^2 - 2x + x^2 + x = 5 \\ x^2 + 2x < 0. \end{cases}$$

   Рекомендуем сначала решить квадратное неравенство $x^2 + 2x > 0$.

   Ответ: $x_1 = -\frac{5}{2}$, $x_2 = 1$.

3. ***Иррациональное уравнение***: $\sqrt{x+2} = x$.

   Уравнение равносильно системе $\begin{cases} x + 2 = x^2, \\ x \geq 0. \end{cases}$

   Заметим, что указывать в ОДЗ $x + 2 \geq 0$ нет надобности, так как всякое решение уравнения, полученного после возведения в квадрат, автоматически попадет в ОДЗ. Ведь если верно, что $x + 2 = x^2$, то $x + 2 \geq 0$, так как $x^2 \geq 0$. Наоборот, пропуск условия $x \geq 0$ нарушает равносильность.

   Ответ: $x = 2$.

4. ***Показательное уравнение***: $2^{x+1} + 2^{2-x} = 9$.

Замена $2^x = y$ немедленно приводит его к алгебраическому: $2y + \frac{4}{y} = 9$.

Ответ: $x_1 = -1, x_2 = 2$.

5. *Логарифмическое уравнение*: $\log_2(3x - x^2) = 1 - \log_2(x - 1)$.
   При потенцировании теряется информация об ОДЗ. Поэтому выпишем ОДЗ в явном виде: $\begin{cases} 3x - x^2 > 0, \\ x - 1 > 0. \end{cases}$
   Решением этой системы неравенств будет интервал (1; 3). Теперь потенцируем, перенося логарифм в левую часть:
   $(3x - x^2)(x - 1) = 2 \Leftrightarrow x^3 - 4x^2 + 3x + 2 = 0$.
   Подобрав один корень $x = 2$, выделяем множитель $(x - 2)$:
   $x^3 - 4x^2 + 3x + 2 = x^3 - 2x^2 - 2x^2 + 4x - x + 2 = (x - 2)(x^2 - 2x - 1)$.
   Корни квадратного множителя: $x = 1 \pm \sqrt{2}$. Сопоставляя с ОДЗ, получаем *ответ*: $x_1 = 2, x_2 = 1 + \sqrt{2}$.
6. *Тригонометрическое уравнение*: $3\sin x + 4\cos x = \frac{24}{5}$.
   Делаем замену $\sin x = \frac{2t}{1+t^2}$, $\cos x = \frac{1-t^2}{1+t^2}$, где $t = \operatorname{tg}\frac{x}{2}$, и получаем уравнение

$$\frac{3t}{1+t^2} + \frac{2(1-t^2)}{1+t^2} = \frac{12}{5} \Leftrightarrow \left[\begin{array}{l} t_1 = \frac{1}{2} \\ t_2 = \frac{2}{11} \end{array}\right. \Leftrightarrow \begin{array}{l} x = 2\operatorname{arctg}\frac{1}{2} + 2\pi n \\ x = 2\operatorname{arctg}\frac{2}{11} + 2\pi n \end{array} \quad (n \in Z).$$

Приближенные методы вычисления корней

Во многих случаях при решении уравнений их корни находят приближенно. Для этого в математике накоплены различные методы приближенных вычислений. Обычно они дают последовательность приближений к искомому числу. Примером может служить способ извлечения квадратного корня, знакомый из курса алгебры.

Простейшим методом приближенного вычисления корней является метод половинного деления. Допустим, что известен промежуток $[a; b]$, на котором лежит искомый корень. Приближенно строится график функции f на этом промежутке (например, так, как это изображено на рис. 400).

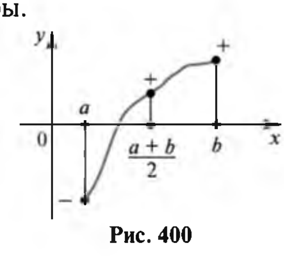


Рис. 400

Вычисляя $f(a)$ и $f(b)$, видим, что эти числа разных знаков: $f(a) < 0$, $f(b) > 0$. Вычисляем значение функции f в середине отрезка $[a; b]$. Из двух половин отрезка $[a; b]$ берем ту, на концах которой знаки функции различны. Очевидно, корень x лежит внутри нового отрезка. Совершаем с ним ту же процедуру: делим его пополам, вычисляем значение функции f в точке деления и берем ту половину отрезка, на концах которой знаки функции f различны. Так мы получим последовательность отрезков, длина которых убывает и внутри которых лежит искомый корень. Это и означает, что получена последовательность приближенных значений искомого корня.

И. Ньютону принадлежит идея *метода касательных*. Об этом способе приближенного вычисления корней можно получить представление, рассматривая рис. 401. Приближенные значения корня получаются построением касательных к графику функции. Уравнение касательной написать нетрудно, а затем найти точку ее пересечения с осью x, что и дает приближенное значение корня функции. Вместо касательных можно проводить хорды (рис. 402) и поступать аналогично (*метод хорд*).

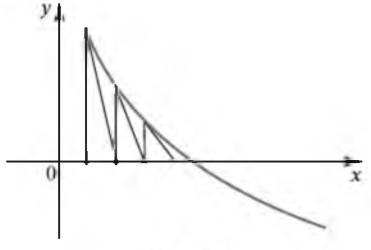


Рис. 401

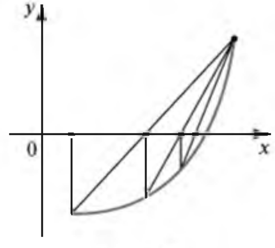


Рис. 402

Контрольные вопросы и задания

1. Сколько корней имеет линейное уравнение $ax + b = 0$?
2. Сколько корней имеет квадратное уравнение $x^2 + px + q = 0$?
3. Сколько корней имеет уравнение $x^n = a$ при различных n и a (n — натуральное число)?
4. Сколько корней имеет простейшее показательное уравнение $a^x > b$ при различных a и b?
5. При каких значениях a уравнение $\sin x = a$ не имеет решений?
6. В чем состоит графический метод решения уравнений?
7. Какие приближенные методы нахождения корней уравнения вы знаете?

8. Почему при решении уравнения вида $f(x) = 0$ стараются разложить на множители левую часть?
9. Придумайте тригонометрическое уравнение, которое с помощью введения нового неизвестного сводилось бы к квадратному.
10. Придумайте показательное уравнение, которое с помощью замены неизвестного сводилось бы к линейному.

§ 43. Неравенства с одним неизвестным

Общие приемы

Решение неравенств (так же как и решение уравнений) обычно распадается на два шага — преобразование неравенства к одному из стандартных и решение стандартного неравенства. К стандартным неравенствам мы отнесем следующие типы неравенств, изученные нами ранее (из возможных четырех знаков неравенства мы выбираем одно):

1. Линейное неравенство $ax + b > 0$.
2. Квадратное неравенство $ax^2 + bx + c > 0$.
3. Степенное неравенство $x^n > a$.
4. Показательное неравенство $a^x > b$.
5. Логарифмическое неравенство $\log_a x > b$.

Общие приемы решения уравнений и неравенств аналогичны. Так же как и для уравнения, при решении неравенств помогает разложение на множители. Решение неравенства вида $\square \cdot \triangle > 0$ можно заменить решением двух систем неравенств:

$$\begin{cases} \square > 0, \\ \triangle > 0 \end{cases} \text{ и } \begin{cases} \square > 0, \\ \triangle > 0. \end{cases}$$

В то же время если множители $\square$ или $\triangle$ являются линейными или произведениями линейных, то не стоит сводить решение неравенства к системе — проще применить метод интервалов, который сильно сокращает количество вычислений.

Важнейшим методом решения неравенств является метод замены неизвестного.

Пример

Решить неравенство
$|2^{2x+3} - 2^{x+5} + 19| \leq 5.$

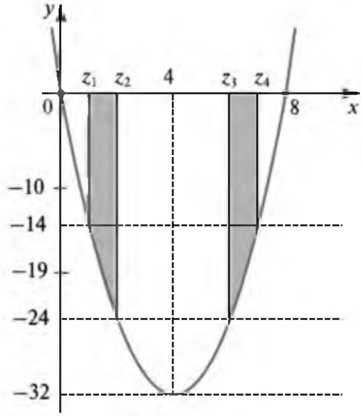


Рис. 403

Прежде всего, сделаем замену $2^{2x+2} = z^2$, тогда $2^{x+1} = z$ и неравенство примет вид $|2z^2 - 16z + 19| \le 5$. Изобразим график квадратного трехчлена $y = 2z^2 - 16z$ (рис. 403). Решением неравенства $|y + 19| \le 5$, как видно из графика, является объединение двух отрезков: $[z_1; z_2]$ и $[z_3; z_4]$, где z_1, z_4 — решения уравнения $y = -14$, а z_2, z_3 — решения уравнения $y = -24$. Решая эти уравнения, находим: $z_1 = 1$, $z_2 = 2$, $z_3 = 6$, $z_4 = 7$. Учитывая, что функция является возрастающей, решаем стандартные неравенства и записываем *ответ*: $[-1; 0] \cup [\log_2 6 - 1; \log_2 7 - 1]$.

Примеры решения неравенств

1. *Алгебраическое неравенство*: $\frac{1}{x+1} + \frac{2}{x+3} < \frac{3}{x+2}$.

Перенесем правую часть влево, приведем к общему знаменателю и разложим на множители числитель дроби:

$$\frac{1}{x+1} + \frac{2}{x+3} - \frac{3}{x+2} < 0 \Leftrightarrow \frac{x^2+5x+6+2(x^2+3x+2)-3(x^2+4x+3)}{(x+1)(x+2)(x+3)} < 0 \Leftrightarrow$$

$$\Leftrightarrow \frac{-x+1}{(x+1)(x+2)(x+3)} < 0 \Leftrightarrow \frac{x-1}{(x+1)(x+2)(x+3)} > 0.$$

Применяем метод интервалов, с помощью числовой оси (рис. 404) решаем неравенство и получаем ответ: $x < -3$, $-2 < x < -1$, $x > 1$.

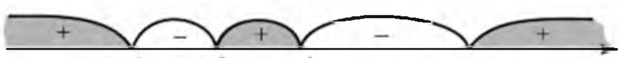


Рис. 404

2. *Иррациональное неравенство*: $\sqrt{x+2} > x$.

ОДЗ: $x + 2 \ge 0 \Leftrightarrow x \ge -2$.

Если иррациональное уравнение мы смело возводили в квадрат, так как всегда можно было проверить нарушение равносильности, подставляя корни полученного уравнения, то при решении неравенства нужно поступать аккуратнее. Заметим, что неравенство $a > b$, где $a \geq 0$, $b < 0$, является всегда верным, какие бы значения указанных знаков ни подставляли вместо a и b. Поэтому если $x < 0$, то неравенство $\sqrt{x+2} > x$ будет верным. Итак, все отрицательные числа, входящие в ОДЗ, будут решениями неравенства. Нанесем их на числовую ось. Пусть $x \geq 0$. Возведение в квадрат теперь не нарушает равносильности: $x + 2 > x^2 \Leftrightarrow$ $\Leftrightarrow x^2 - x - 2 < 0$.

Корни квадратного трехчлена $x_1 = -1$, $x_2 = 2$ наносим на числовую ось; решением будут числа $0 \leq x < 2$. *Ответ*: $-2 \leq x < 2$.

3. *Логарифмическое неравенство*: $\log_{\frac{1}{2}} \frac{x}{x^2 - x - 2} \geq \frac{1}{\lg 2} - 2$.

Сначала преобразуем правую часть: $\frac{1}{\lg 2} - 2 = \log_2 10 - 2 =$ $= -\log_{\frac{1}{2}} 10 - \log_{\frac{1}{2}} \frac{1}{4} = \log_{\frac{1}{2}} \frac{2}{5}$. Стандартное логарифмическое неравенство $\log_{\frac{1}{2}} \square \geq \log_{\frac{1}{2}} \frac{2}{5}$ равносильно системе $\begin{cases} \square \leq \frac{2}{5}, \\ \square > 0. \end{cases}$ Решаем каждое из неравенств системы методом интервалов, предварительно сделав преобразования: $\frac{x}{x^2 - x - 2} - \frac{2}{5} = \frac{-(2x^2 - 7x - 4)}{5(x+1)(x-2)}$.

Корни числителя: $x_1 = -\frac{1}{2}$, $x_2 = 4$. Решение системы неравенств изображено на рис. 405. *Ответ*: $-\frac{1}{2} \leq x < 0$; $x \geq 4$.

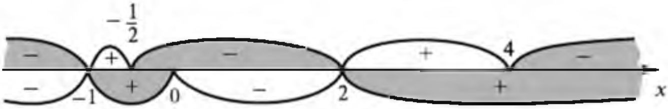


Рис. 405

Контрольные вопросы и задания

1. Каким может быть множество решений линейного неравенства?
2. Перечислите все возможные типы ответов при решении квадратного неравенства $ax^2 + bx + c \leq 0$.

3. Какой вид может иметь множество решений неравенства $x^2 \le a$?
4. При всех ли a и b неравенство $ax > b$ имеет хотя бы одно решение?
5. Может ли логарифмическое неравенство $\log_a x < b$, быть верным при любом положительном значении x?
6. Можно ли при решении неравенства умножить обе его части на $x^2 + 1$?
7. Сформулируйте, что происходит с неравенством при умножении обеих его частей на выражение $f(x)$.
8. Приведите пример логарифмического неравенства, не имеющего решений.
9. Левая часть неравенства является произведением трех линейных множителей (в правой части — нуль). Подсчитайте, сколько случаев вам придется рассмотреть, перебирая все комбинации знаков множителей, и сколько случаев — при использовании метода интервалов.
10. Приведите пример неравенства, множество решений которого состоит из двух чисел.

§ 44. Системы уравнений

Способ подстановки

Системы уравнений появляются при решении задач, где неизвестной является не одна величина, а несколько. Эти величины связаны определенными зависимостями, которые записываются в виде уравнений.

Если система имеет хотя бы одно решение, она называется совместной. Если решений у системы нет, она называется несовместной.

Слово «несовместность» наглядно показывает, что разные уравнения системы накладывают несовместимые друг с другом условия, которым должны удовлетворять неизвестные.

Одним из основных способов решения систем является способ подстановки. Рассмотрим, например, систему двух уравнений с двумя неизвестными x и y. Часто удается одно уравнение преобразовать так, чтобы одно неизвестное явно выражалось как функция другого. Тогда, подставляя его во второе уравнение, мы получим уравнение с одним неизвестным.

Примеры

1. $\begin{cases} x^2+y^2=25 \\ x-y=1 \end{cases}$

2. $\begin{cases} x^2+y^2=5 \\ x^2+y=3 \end{cases}$

3. $\begin{cases} \dfrac{1}{x}+\dfrac{3}{y}=2 \\ xy=3 \end{cases}$

4. $\begin{cases} 2x+3y=4xy \\ x+y=\dfrac{3}{2}xy \end{cases}$

В каждой из четырех систем второе уравнение системы можно решить относительно y, т.е. преобразовать его к виду $y = f(x)$, а именно:

1. $x - y = 1 \Leftrightarrow y = x - 1.$
2. $x^2 + y = 3 \Leftrightarrow y = 3 - x^2.$
3. $xy = 3 \Leftrightarrow y = \dfrac{3}{x}.$
4. $x + y = \dfrac{3}{2}xy \Leftrightarrow y\left(1-\dfrac{3}{2}x\right) = -x \Leftrightarrow y = \dfrac{x}{\dfrac{3}{2}x-1}.$

Подставляя $y = f(x)$ в первое уравнение системы, получаем уравнение с одним неизвестным:

1. $x^2 + (x - 1)^2 = 25.$
2. $x^2 + (3 - x^2)^2 = 5.$
3. $\dfrac{1}{x} + x = 2.$
4. $2x + \dfrac{3x}{\dfrac{3}{2}x-1} = 4x\dfrac{x}{\dfrac{3}{2}x-1}.$

Решая получившиеся уравнения с одним неизвестным, находим его корни — значения неизвестного, а затем для каждого из них находим соответствующее значение по формуле $y = f(x)$:

1. $x^2+(x-1)^2=25 \Leftrightarrow 2x^2-2x-24=0 \Leftrightarrow x^2-x-12=0 \Leftrightarrow x_1=4, x_2=-3,$
   $y_1 = x_1 - 1 = 3,\ y_2 = -3 - 1 = -4.$
   Решения системы: $\begin{cases} x_1=4, \\ y_1=3 \end{cases}$ и $\begin{cases} x_2=-3, \\ y_2=-4. \end{cases}$
2. $x^2 + (3 - x^2)^2 = 5 \Leftrightarrow x^4 - 5x^2 + 4 = 0,\ (x^2)_1 = 1,\ (x^2)_2 = 4.$
   Уравнение имеет четыре корня, а система — четыре решения:
   $\begin{cases} x_1=1, \\ y_1=2; \end{cases}$ $\begin{cases} x_2=-1, \\ y_2=2; \end{cases}$ $\begin{cases} x_3=2, \\ y_3=-1; \end{cases}$ $\begin{cases} x_4=-2, \\ y_4=-1. \end{cases}$

3. $x + \dfrac{1}{x} = 2 \Leftrightarrow x^2 - 2x + 1 = 0 \Leftrightarrow x = 1.$

Решение одно: $x = 1$, $y = 3$.

4. $2x + \dfrac{3x}{\frac{3}{2}x - 1} = 4x\dfrac{x}{\frac{3}{2}x-1} \quad \Leftrightarrow \quad 2x + \dfrac{2\cdot 3x}{3x-2} = \dfrac{4x\cdot 2x}{3x-2} \Leftrightarrow x\left(1 + \dfrac{3}{3x-2}\right) =$

$= x\dfrac{4x}{3x-2} \Leftrightarrow x = 0$ или $1 + \dfrac{3}{3x-2} = \dfrac{4x}{3x-2} \Leftrightarrow 3x - 2 + 3 = 4x \Leftrightarrow$
$\Leftrightarrow x = 1.$

Ответ: $\begin{cases} x_1 = 0, \\ y_1 = 0; \end{cases} \begin{cases} x_2 = 1, \\ y_2 = 2. \end{cases}$

Способ подстановки возможен не всегда, а кроме того, не всегда выгоден и тогда, когда возможен. Часто из уравнений системы удается получить новое уравнение — их следствие — более простого вида. Так, в четвертом из рассматривающихся выше примеров можно исключить произведение xy, стоящее справа:

$$\begin{cases} 2x + 3y = 4xy \\ x + y = \dfrac{3}{2}xy \end{cases} \Leftrightarrow \begin{cases} \dfrac{2x+3y}{4} = xy, \\ \dfrac{2}{3}(x+y) = xy. \end{cases} \quad \dfrac{2x+3y}{4} = \dfrac{2}{3}(x+y),$$

Последнее соотношение является линейным. И из него легче находится связь между x и y: $y = 2x$.

Важным приемом, часто позволяющим упростить систему, является замена неизвестных. Так, во втором примере полезно заменить x^2 на z и получить более простую систему:

$$\begin{cases} z + y^2 = 5 \\ z + y = 3 \end{cases}.$$

Системы с двумя неизвестными и их решения можно изобразить графически на координатной плоскости. На рисунках 406—409 показаны кривые уравнений для написанных выше систем. Точки пересечения кривых (а точнее, их координаты) — решения систем.

Есть некоторые типы систем, для которых известны стандартные методы решения. Рассмотрим два из них: симметричные системы и линейные системы.

Симметричные системы

Симметричными называются системы, составленные из выражений, являющихся симметричными относительно всех неизвестных.

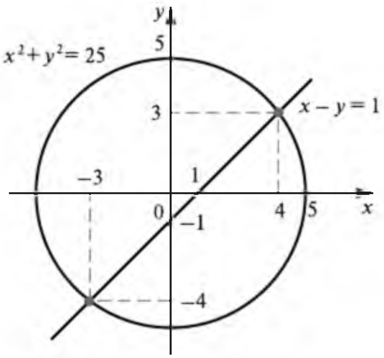


Рис. 406

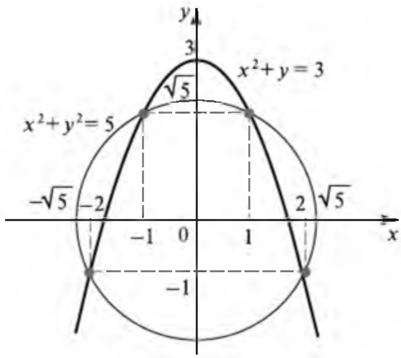


Рис. 407

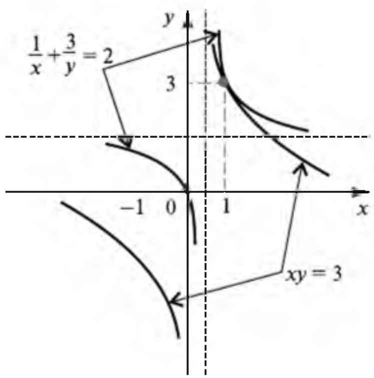


Рис. 408

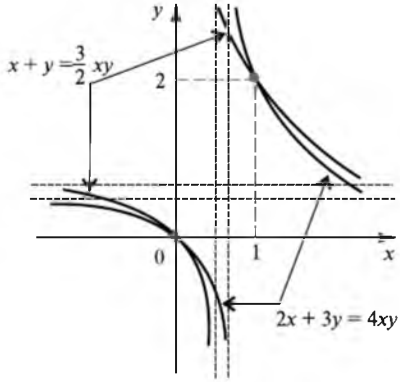


Рис. 409

Приведем примеры различных симметричных выражений для двух неизвестных x и y:

1. $u_1 = x + y$.
2. $u_2 = xy$.
3. $u_3 = x^3 + y^3$.
4. $u_4 = \dfrac{1}{x} + \dfrac{1}{y}$.
5. $u_5 = x^2 + y^2$.
6. $u_6 = \dfrac{x-1}{y} + \dfrac{y-1}{x}$.

Решение простейшей симметричной системы $\begin{cases} x + y = \alpha, \\ xy = \beta, \end{cases}$ основано на теореме, обратной теореме Виета: x и y, удовлетворяющие указанной системе, являются корнями квадратного уравнения $t^2 - \alpha t + \beta = 0$. Этот вывод можно получить, подставив из первого уравнения во второе $y = \alpha - x$.

Итак, для решения простейшей симметричной системы надо составить квадратное уравнение с заданными суммой и произведением корней и решить его. Найденные корни будут значениями x и y.

Примеры

1. $\begin{cases} x+y=3, \\ xy=-4. \end{cases}$

Составляем квадратное уравнение: $t^2 - 3t - 4 = 0$, откуда $t_1 = 4$, $t_2 = -1$. Решения систем: $\begin{cases} x_1 = 4, \\ y_1 = -1 \end{cases}$ и $\begin{cases} x_2 = -1, \\ y_2 = 4. \end{cases}$

Решение других симметричных систем основано на том, что всякое симметричное выражение относительно x и y можно представить через $u = x + y$ и $v = xy$. Например,

$x^2 + y^2 = x^2 + 2xy + y^2 - 2xy = (x + y)^2 - 2xy = u^2 - 2v$, $\frac{1}{x} + \frac{1}{y} = \frac{y+x}{xy} = \frac{u}{v}$,

$x^3 + y^3 = (x + y)(x^2 + y^2 - xy) = u(u^2 - 2v - v) = u(u^2 - 3v)$,

$\frac{x-1}{y} + \frac{y-1}{x} = \frac{x^2+y^2-x-y}{xy} = \frac{u^2-2v-u}{v}$.

Делая в симметричной системе замену $x + y = u$, $xy = v$, получаем более простую систему относительно u и v, а затем, найдя числовые значения u и v, приходим к решению простейших симметричных систем $\begin{cases} x+y=u, \\ xy=v. \end{cases}$

2. $\begin{cases} x^2+y^2=25, \\ x+y=7; \end{cases}$ $\begin{cases} u^2-2v=25, \\ u=7; \end{cases}$ $\Leftrightarrow$ $\begin{cases} u=7, \\ v=\frac{1}{2}(u^2-25)=12. \end{cases}$

Решаем систему $\begin{cases} x+y=7, \\ xy=12. \end{cases}$

3. $\begin{cases} x^3+y^3=7, \\ x^2+3xy+y^2=-1. \end{cases}$

Воспользуемся найденным выше выражением $x^3 + y^3$ через u и v:

$$\begin{cases} u(u^2-3v)=7, \\ u^2+v=-1. \end{cases}$$

Из второго уравнения находим $v = -u^2 - 1$. Тогда $u^3 + 3u(u^2 + 1) = 7 \Leftrightarrow 4u^3 + 3u - 7 = 0 \Leftrightarrow (u - 1)(4u^2 + 4u + 7) = 0$, $u_1 = 1$, $v = -u^2 - 1 = -2$ и $4u^2 + 4u + 7 = 0$ — корней нет. Решаем систему:

$$\begin{cases} x+y=1, \\ xy=2. \end{cases}$$

Линейные системы

Раньше вы изучали системы двух линейных уравнений с двумя неизвестными вида

$$\begin{cases} ax + by = k, \\ cx + dy = l. \end{cases}$$

На практике встречаются системы линейных уравнений с большим количеством неизвестных. Так, в задачах математической экономики существуют системы, состоящие из нескольких сотен уравнений с таким же примерно числом неизвестных. Для их решения разработаны мощные машинные методы. Эти методы в основном имитируют знакомый вам метод подстановки, которым, в принципе, можно решить любую такую систему. Основную роль при этом играют компактные способы записи систем и их преобразований. Представьте только себе — система из тысячи уравнений с тысячью неизвестными содержит миллион коэффициентов.

Рассмотрим более простой пример — систему трех линейных уравнений с тремя неизвестными.

Пример

$$\begin{cases} x + 2y + 3z = 2, \\ 2x + 3y - 4z = -5, \\ 3x + y + z = 3. \end{cases}$$

Будем решать систему методом исключения неизвестных. Чтобы исключить x из второго и третьего уравнений, надо вычесть из них первое, умноженное соответственно на 2 и на 3:

$$\begin{cases} x + 2y + 3z = 2, \\ - y - 10z = -9, \\ -5y - 8z = -3. \end{cases}$$

Удобно умножить второе и третье уравнение на -1, а затем из третьего уравнения вычесть второе, умноженное на 5. Получим «треугольную» систему

$$\begin{cases} x + 2y + 3z = 2, \\ y + 10z = 9, \\ 42z = 42. \end{cases}$$

Из третьего уравнения системы находим $z = 1$. Подставляя это значение во второе уравнение, находим $y = 9 - 10 = -1$.
Подставляя $z = 1$, $y = -1$ в первое уравнение, находим $x = 1$.
Ответ: $x = 1$, $y = 1$, $z = 1$.

Показанный на этом примере способ решения линейной системы называется *методом Гаусса*, по имени великого немецкого математика, жившего в первой половине XIX в. Метод Гаусса с различными модификациями используется при решении линейных систем с помощью вычислительных машин.

Контрольные вопросы и задания

1. Какая система уравнений называется совместной?
2. Приведите пример несовместной системы.
3. Что такое решение системы уравнений?
4. Как графически изобразить решение системы двух уравнений с двумя неизвестными?
5. Каким методом решают линейные системы уравнений?
6. В чем состоит теорема, обратная теореме Виета?
7. Как выражается сумма квадратов чисел через их сумму и произведение?
8. Приведите примеры систем выражений, симметричных относительно x и y.
9. Как симметричность выражения $f(x, y)$ скажется на графике зависимости $f(x, y) = 0$? Приведите примеры.
10. В чем состоит метод подстановки?

Заключительная беседа

Тождества

Мы определили тождество как равенство двух выражений, справедливое при всех допустимых наборах значений букв, входящих в эти выражения. Такая точка зрения свойственна теории функций — мы рассматриваем две части равенства как функции и называем эти части тождественно равными, если они совпадают как функции, т.е. если они при одних и тех же значениях аргумента принимают равные значения. Возможна другая точка зрения на тождества, которая более тесно связана с алгеброй.

В алгебре многочлен рассматривается не как функция, а как некоторое формальное выражение, составленное из одночленов. Мы умеем совершать различные операции над многочленами, не задумываясь при этом над тем, какие значения можно подставлять в многочлен вместо букв. В алгебре два многочлена равны, если после приведения подобных членов окажется, что они составлены из одинаковых одночленов, т.е. если выполняется формальное, почленное равенство. Так, проверяя тождество $a^3 - b^3 = (a - b)(a^2 + ab + b^2)$, мы совсем не занимаемся подстановкой в обе части значений a и b (тем более, что неясно, сколько их надо подставлять), а преобразуем правую часть и убеждаемся, что она формально совпадает с левой.

Проверке формального совпадения многочленов может помочь их запись, принятая в качестве стандартной. Например, если многочлены от одной буквы x записывать по убывающим степеням (как мы привыкли), то тождество многочленов означает равенство их степеней и совпадение коэффициентов, стоящих на одинаковых местах.

Возникает естественный вопрос: как связаны между собой функциональное и алгебраическое определения тождества? Разумеется, если два многочлена равны формально, то они принимают одинаковые значения при всех значениях букв. Обратное заключение составляет содержание трудной теоремы алгебры — теоремы о тождестве. Поясним смысл этой теоремы для простейшего случая многочленов от одной буквы x.

Прежде всего, заметим, что от равенства $f(x) = g(x)$ всегда можно перейти к равенству $f(x) - g(x) = 0$, как бы мы ни определяли понятие тождества. Это означает, что теорему о тождестве можно доказывать в таком упрощенном варианте:

если многочлен $F(x)$ при всяком значении x равен нулю, то этот многочлен нулевой.

т.е. не содержит ни одного ненулевого одночлена. Если многочлен $F(x)$ имеет степень n, то, оказывается, достаточно проверить, что он равен нулю при $(n + 1)$ значении x. Тогда этот многочлен нулевой. Иными словами:

если многочлен степени n имеет $(n + 1)$ корень, то этот многочлен нулевой.

В такой формулировке теорема допускает уже не очень сложное доказательство.

Итак, полезно запомнить, что ненулевой многочлен не может иметь корней больше, чем его степень. Возможна другая формулировка:

если два многочлена степени n совпадают в $(n+1)$-й точке, то эти многочлены формально равны.

Эта формулировка очень полезна при доказательстве различных тождеств.

В применении к многочленам первой степени нам знакома геометрическая формулировка этой теоремы: через две точки проходит только одна прямая. Аналогично, для совпадения двух квадратных трехчленов достаточно равенства их значений в трех точках.

Кроме равенства многочленов можно определить равенство дробей с алгебраической точки зрения: две дроби $\frac{f_1(x)}{g_1(x)}$ и $\frac{f_2(x)}{g_2(x)}$ считаются равными, если формально равны многочлены $f_1(x)\ g_2(x)$ и $g_1(x)\,f_2(x)$.

В более усложненном варианте алгебраический подход возможен и к тригонометрическим тождествам. Так, тождествам, содержащим степени $\sin x$ и $\cos x$, можно придать условный характер: доказать тождество, используя из тригонометрии лишь соотношение $\sin^2 x + \cos^2 x = 1$. Такую задачу можно решить, делая лишь алгебраические преобразования и не вспоминая о том, что такое синус и косинус. Приведем пример условного тождества в алгебре:

$$\text{если } a + b + c = 0, \text{ то } a^3 + b^3 + c^3 = 3abc.$$

Доказательство неравенств

Наряду с тождествами-равенствами, выполняющимися тождественно, — существуют тождественно выполняющиеся неравенства, т.е. неравенства, верные при любых допустимых значениях входящих в них букв.

Примеры

1. $x^2 \geq 0$.
2. $a^2 + b^2 + c^2 \geq 0$, причем равенство нулю возможно лишь при $a = b = c = 0$.
3. $x^2 + px + q > 0$, если $p^2 - 4q < 0$.

Задачи на доказательство неравенств (т.е. на доказательство того, что неравенство выполняется при всех допустимых значениях букв) решаются с помощью цепочки преобразований, приводящей к равносильному известному неравенству.

Пример

Доказать неравенство $\frac{a+b}{2} \ge \sqrt{ab}$, где $a \ge 0$, $b \ge 0$.

Делаем цепочку преобразований:

$$\frac{a+b}{2} \ge \sqrt{ab} \Leftrightarrow a+b \ge 2\sqrt{ab} \Leftrightarrow a-2\sqrt{ab}+b \ge 0 \Leftrightarrow (\sqrt{a}-\sqrt{b})^2 \ge 0.$$

Последнее неравенство всегда верно; следовательно, всегда верно исходное.

Полученное неравенство (его называют неравенством о среднем арифметическом и среднем геометрическом двух чисел) можно применять к доказательству других неравенств. Убедитесь, например, что следующие неравенства являются следствиями доказанного:

1. $a+\frac{1}{a} \ge 2$ при $a > 0$.
2. Если $ab = 1$, то $a + b \ge 2$ (где $a > 0$, $b > 0$).
3. Если $a_1a_2 = 1$, то $(1 + a_1)(1 + a_2) \ge 4$ (где $a_1 > 0$, $a_2 > 0$).
4. $\frac{a^4+b^4+c^4+d^4}{4} \ge abcd$, где ($a \ge 0$, $b \ge 0$, $c \ge 0$, $d \ge 0$).

Использование производной дает мощный способ доказательства неравенств с одной переменной. Этот способ основан на следующем соображении:

если в точке x_0 выполняется условие $f(x_0) \ge 0$ и для всех $x \ge x_0$ выполняется условие $f'(x_0) \ge 0$, то для всех $x > x_0$, верно неравенство $f(x) \ge 0$.

(Разберитесь в справедливости сформулированного правила.)

Пример (неравенство Бернулли)

$(1 + x)^k \ge 1 + kx$ (где $x \ge 0$, $k \ge 1$).

Для доказательства рассмотрим функцию $y = f(x)$, где $f(x) = (1 + x)^k - 1 - kx$. Имеем $f(0) = 0$, $f'(x) = k(1 + x)^{k-1} - k = k((1 + x)^{k-1} - 1)$. Так как $x \ge 0$, $k \ge 1$, то $(1 + x)^{k-1} \ge 1$ и $f'(x) \ge 0$. Значит, при $x \ge 0$ функция f возрастает и при всяком $x \ge 0$ имеем $f(x) \ge f(0)$, что и требовалось доказать.

Алгебраические уравнения

Алгебраическое уравнение — это уравнение вида

$$xn + an_{-1}xn^{-1} + \ldots + a_1x + a_0 = 0.$$

Число n называется степенью уравнения.

Уравнение первой степени (или линейное уравнение) решается с помощью арифметических операций. Формула для решения уравнения второй степени (или квадратного уравнения) известна с глубокой древности. В нее входит операция извлечения квадратного корня. Решение уравнения произвольной степени в течение многих веков считалось основной задачей алгебры.

Постановка вопроса о решении алгебраического уравнения может быть различной. Почему «не решается» данное нам уравнение? Рассмотрим возможные ответы на этот вопрос.

1. Нам «не хватает» имеющихся чисел. Уравнение $x^2 + 2x + 5 = 0$ не имеет вещественных корней. Можно, конечно, на этом утверждении остановиться. Однако полезно, как это было сделано еще в XVI в., ввести комплексные числа, с которыми вы немного знакомы. Комплексное число имеет вид $a + bi$, где a и b — вещественные числа, а символ i (мнимая единица) обозначает такое число, для которого $i^2 = -1$. Комплексные числа $x_1 = -1 - 2i$ и $x_2 = -1 + 2i$ являются корнями написанного выше квадратного уравнения.

Если мы разрешим числу x принимать не только вещественные, но и комплексные значения, то отпадет вопрос о существовании корня алгебраического уравнения. В 1799 году Гаусс доказал замечательную теорему, которую часто называют *основной теоремой алгебры*:

всякое алгебраическое уравнение имеет хотя бы один комплексный корень.

2. Мы не можем разложить левую часть уравнения на множители. Возьмем, например, уравнение $x^5 + x + 1 = 0$. Не сразу бросается в глаза, что левую часть можно разложить на множители:

$$x^5 + x + 1 = x^5 - x^2 + x^2 + x + 1 = x^2(x^3 - 1) + x^2 + x + 1 =$$
$$= (x^2 + x + 1)(x^3 - x^2 + 1).$$

После разложения на множители получим уравнения меньших степеней: $x^2 + x + 1 = 0$ и $x^3 - x^2 + 1 = 0$. Однако этот прием проходит далеко не всегда. Так, многочлен $x^2 - x + 1$ уже нельзя разложить на множители с целыми коэффициентами. Известен алгоритм, который позволяет разложить любой многочлен с целыми коэффициентами на множители с целыми коэффициентами, если это возможно. Частный случай применения этого алгоритма мы неоднократно использовали: если многочлен $x^n + a_{n-1}x^{n-1} + \ldots + a_0$ с целыми коэффициентами имеет множитель вида $x - c$, где c — целое число (являющееся, конечно, корнем многочлена), то свободный член a_0 делится на c. Эта теорема позволяет перебором делителей свободного члена и проверкой найти целые корни многочлена с целыми коэффициентами.

3. Мы не знаем общей формулы для корней уравнения. Простая формула корней квадратного уравнения вызывала желание математиков найти формулы корней уравнения более высокой степени. В XVI в. эта задача была решена для уравнений 3-й и 4-й степеней. Хотя эти формулы громоздки и не употребляются для реального вычисления корней, принципиальное их значение велико: они позволяют записать корни уравнений 3-й (4-й) степени как некоторую функцию от коэффициентов этих уравнений. Эта функция содержит операции извлечения корней 3-й (и 4-й) степени. Долго изучавшийся вопрос о том, существует ли формула, выражающая корни уравнения 5-й степени через его коэффициенты с помощью радикалов, получил отрицательное решение в работах Абеля (1802—1829) и Галуа (1811—1832) в начале XIX в.

Итак, как правило, для алгебраического уравнения высокой степени мы не можем указать общей формулы его корней. Для приближенного вычисления корней используют методы анализа, примеры которых мы рассматривали.

Различные приближенные методы нахождения корней уравнения часто базируются на следующем:

если на концах промежутка функция $y = f(x)$ принимает значения разных знаков, то внутри этого промежутка уравнение $f(x) = 0$ имеет корень.

Это утверждение верно для всех непрерывных функций. С его помощью нетрудно, например, доказать, что всякий многочлен нечетной степени имеет вещественный корень. Например, кубическое уравнение $x^3 + ax^2 + bx + c = 0$ всегда имеет хотя бы одно решение, так как левая часть при больших по модулю и отрицательных x меньше нуля (слагаемое x^3 «перевесит» все остальные), а при положительных больших x станет больше нуля.

Для разрывных функций сформулированное утверждение может оказаться неверным, как показывает простой пример функции $y = \dfrac{1}{x}$, не имеющей корней, но принимающей значения разных знаков.

Контрольные вопросы и задания

1. Какие определения тождества многочленов вы знаете? Чем они отличаются друг от друга?
2. Сколько корней может иметь многочлен степени n?
3. Каким количеством точек определяется многочлен n-й степени?

4. В каком количестве точек достаточно проверить совпадение значений двух многочленов четвертой степени, чтобы доказать их тождественное равенство?
5. Приведите примеры тождественно выполняющихся неравенств.
6. В чем состоит неравенство для среднего арифметического и среднего геометрического двух чисел? Предложите его обобщение для n чисел.
7. Как с помощью признака монотонности функции можно доказывать неравенства с одной переменной?
8. В чем состоит основная теорема алгебры?
9. Начиная с какого n не существует общей формулы решения уравнения n-й степени?
10. Может ли многочлен пятой степени не иметь вещественных корней?