АКАДЕМИЯ НАУК СССР
МАТЕМАТИЧЕСКИЙ ИНСТИТУТ им. В. А. СТЕКЛОВА

МАТЕМАТИКА
ЕЕ СОДЕРЖАНИЕ, МЕТОДЫ И ЗНАЧЕНИЕ

ТОМ ПЕРВЫЙ

ИЗДАТЕЛЬСТВО АКАДЕМИИ НАУК СССР
МОСКВА 1956

ПРЕДИСЛОВИЕ

Возникшая еще в древности из практических потребностей, математика выросла в громадную систему разветвленных дисциплин. Как и другие науки, она отражает законы материальной действительности и служит могучим орудием познания и покорения природы. Но свойственный математике высокий уровень абстракции делает новые ее разделы сравнительно мало доступными для неспециалиста. Тот же отвлеченный характер математики порождал еще в древности идеалистические представления о ее независимости от материальной действительности.

Коллектив авторов при составлении этой книги исходил из намерения ознакомить достаточно широкие круги советской интеллигенции с содержанием и методами отдельных математических дисциплин, их материальными основами и путями развития.

В качестве минимума предварительных математических знаний читателя предполагается знание только курса средней школы, однако в отношении доступности материала каждый из трех томов не является однородным. Желающие впервые познакомиться с началами высшей математики, с пользой прочтут несколько первых глав, но для полного понимания следующих глав необходимо изучение соответствующих учебников. В полном объеме книга окажется доступной в основном лишь читателям, уже имеющим некоторые навыки в применении методов математического анализа (дифференциального и интегрального исчисления). Для таких читателей — представителей естественнонаучных и инженерных специальностей, учителей математики — особенно существенными окажутся главы, вводящие их в более новые разделы математики.

Естественно, что в рамках одной книги нельзя исчерпать всего богатства даже основных направлений математических исследований; некоторая свобода в выборе материала при этом необходима. Но в самых общих чертах эта книга должна дать представление о современном состоянии математики, ее происхождении и перспективах развития в целом. Поэтому книга в известной мере рассчитана и на лиц, владеющих основной частью использованного в ней фактического материала. Она должна способствовать устранению некоторой узости перспективы, свойственной иногда некоторым нашим молодым математикам.

Отдельные главы этой книги написаны разными авторами, их фамилии приведены в оглавлении. Однако как целое книга — результат коллективного труда. Ее общий план, отбор материала, варианты текста отдельных глав подвергались коллективному обсуждению и улучшались на основе живого обмена мнениями. Математики многих городов Советского Союза высказали на организованном институтом обсуждении ценные замечания по первоначальному варианту текста. Эти замечания и предложения были учтены авторами.

Некоторые из авторов принимали также непосредственное участие в подготовке окончательного текста других глав: вводная часть главы II написана в основном Б. Н. Делоне; Д. К. Фаддеев принимал активное участие в подготовке глав IV и XX.

В работе участвовал также ряд лиц, не являющихся авторами отдельных глав: Л. В. Канторовичем написан § 4 главы XIV, О. А. Ладыженской написан § 6 главы VI, А. Г. Постниковым написан § 5 главы X, О. А. Олейник участвовала в подготовке текста главы V, Ю. В. Прохоров участвовал в окончательном редактировании текста главы XI.

В. А. Залгаллер написал некоторые разделы глав I, II, VII и XVII. Редактирование окончательного текста осуществлено В. А. Залгаллером и В. С. Виденским при участии Т. В. Рогозкиной и А. П. Леоновой.

Основная часть иллюстраций выполнена Е. П. Сенькиным.

Редакционная коллегия.

Глава I

ОБЩИЙ ВЗГЛЯД НА МАТЕМАТИКУ

Правильное представление о любой науке не складывается из отдельных, касающихся ее сведений, даже если они довольно обширны. Нужно еще иметь верный взгляд на науку в целом, понимать сущность данной науки. Цель этой главы состоит в том, чтобы дать общее представление о сущности математики. Для этого нет большой необходимости входить в подробное рассмотрение новых математических теорий, потому что уже история этой науки и элементарная математика дают достаточно оснований для общих выводов.

§ 1. ОСОБЕННОСТИ МАТЕМАТИКИ

1. Даже при довольно поверхностном знакомстве с математикой легко заметить характерные ее черты: это, во-первых, ее отвлеченность, во-вторых, точность или, лучше сказать, логическая строгость и как бы непреложность ее выводов и, наконец, чрезвычайная широта ее применений.

Отвлеченность проявляется уже в простом счете. Мы оперируем отвлеченными числами, не заботясь о том, чтобы связывать их каждый раз с конкретными предметами. Мы учим в школе абстрактную таблицу умножения — таблицу умножения чисел вообще, а не числа мальчиков на число яблок или числа яблок на цену яблока и т. п.

Точно так же в геометрии рассматривают, например, прямые линии, а не натянутые нити, причем в понятии геометрической линии отвлекаются от всех свойств, оставляя только протяжение в одном направлении. Вообще понятие о геометрической фигуре является результатом отвлечения от всех свойств реальных предметов, кроме пространственной формы и размеров.

Такого рода отвлечения характерны для всей математики. Понятия о целом числе и о геометрической фигуре — это лишь одни из первоначальных ее понятий. За ними следует едва обозримое множество других, возвышающихся до таких абстракций, как комплексные числа, функции, интегралы, дифференциалы, функционалы, n-мерные и даже бесконечномерные пространства и т. д. и т. п. Абстракции эти как будто громоздятся одна на другую, удаляясь в такую отвлеченность, где, кажется, теряется уже всякая связь с жизнью и где «простой смертный» не поймет ничего, кроме того, что «все это непонятно».

На самом деле это, конечно, не так. И хотя, скажем, понятие n-мерного пространства действительно очень абстрактно, оно тем не менее имеет вполне реальное содержание, понять которое вовсе не так трудно. В этой книге будет, в частности, подчеркнут и пояснен реальный смысл перечисленных абстрактных понятий, и читатель убедится, что все они связаны с жизнью и по своему происхождению и в приложениях.

Впрочем, абстракция — не исключительная принадлежность математики: она свойственна всякой науке, да и всему человеческому мышлению вообще. Поэтому отвлеченность математических понятий не исчерпывает еще особенностей математики.

Математика в отношении своих абстракций отличается еще тем, что она, во-первых, оставляет в них прежде всего количественные отношения и пространственные формы, отвлекаясь от всего остального. Во-вторых, математические абстракции возникают через ряд ступеней; они идут в отвлечении гораздо дальше, чем абстракции, обычные в естественных науках. Эти два момента мы дальше подробно выясним на примерах основных понятий математики: числа и фигуры. Наконец,— и это бросается в глаза,— математика, как таковая, сама по себе вообще почти целиком вращается в кругу абстрактных понятий и их связей. Если естествоиспытатель для доказательства своих утверждений постоянно обращается к опыту, то математик доказывает теоремы только рассуждениями и выкладками.

Конечно, и математики для открытия своих теорем и методов постоянно пользуются моделями, физическими аналогиями, обращаются к множеству отдельных, совершенно конкретных примеров и т. п. Все это служит реальным источником теории, служит для нахождения ее теорем, но каждая теорема окончательно входит в математику только тогда, когда она строго доказана логическим рассуждением. Если бы геометр, докладывая о новой открытой им теореме, стал демонстрировать ее на моделях и этим ограничился, никто из математиков не признал бы теорему доказанной. Требование *доказать* теорему хорошо известно уже из школьного курса геометрии, и оно проходит через всю математику. Мы могли бы измерять углы у оснований тысячи равнобедренных треугольников с огромной точностью, но это не дало бы нам *математического доказательства теоремы* о том, что углы при основании равнобедренного треугольника равны. Математика требует вывести этот результат из основных понятий геометрии. (Теперь при строгом изложении геометрии свойства основных понятий точно формулируют в аксиомах.) И так всегда: доказать теорему для математика означает вывести ее путем рассуждения из начальных свойств, присущих тем понятиям, которые фигурируют в этой теореме. Таким образом, не только понятия, но и метод математики оказывается отвлеченным, умозрительным.

Сами математические выводы отличаются большой логической строгостью. Математическое рассуждение проводится с такой скрупулезностью, которая делает его бесспорным и убедительным для каждого, кто только его поймет. Эта скрупулезность и убедительность математических доказательств хорошо известна уже из курса средней школы. Да и сами математические истины представляются совершенно бесспорными. Недаром говорят: «доказать как дважды два четыре». Здесь математическое соотношение $2 \times 2 = 4$ берется именно как образец неопровержимости и бесспорности.

Однако строгость математики не абсолютна: она развивается; принципы математики не застыли раз навсегда, а движутся и тоже могут служить и служат предметом научных споров.

В конечном счете *источник жизненности математики заключается в том, что ее понятия и выводы при всей своей отвлеченности исходят, как мы убедимся, из действительности и находят широкие применения в других науках, в технике, во всей жизненной практике; это — самое главное для понимания математики.*

Исключительная широта применений математики представляет тоже одну из характерных ее особенностей.

Во-первых, мы постоянно, чуть ли не ежечасно, на производстве, в быту, в общественной жизни пользуемся наиболее распространенными понятиями и выводами математики, вовсе не задумываясь об этом. Так, мы применяем арифметику, считая дни или расходы, а подсчитывая площадь квартиры, используем выводы геометрии. Выводы эти, конечно, очень простые, но полезно вспомнить, что когда-то в древности они были одним из высших достижений зарождавшейся тогда математики.

Во-вторых, вся современная техника была бы невозможна без математики. Без более или менее сложных расчетов не обходится, пожалуй, ни одно техническое усовершенствование; в развитии же новых областей техники математика играет очень важную роль.

Наконец, почти все науки более или менее существенно пользуются математикой. «Точные науки» — механика, астрономия, физика, а также в большой мере и химия — обычно выражают свои законы формулами (как это знакомо каждому еще со школьной скамьи) и развивают свои теории, широко используя математический аппарат. Без математики прогресс этих наук был бы просто невозможен. Поэтому как раз потребности механики, астрономии и физики всегда оказывали прямое, решающее воздействие на развитие математики.

В других науках математика играет меньшую роль, но и там она находит важные применения. Конечно, в изучении таких сложных явлений, как явления биологические и общественные, математический метод по существу не может играть такой же роли, как, скажем, в физике. Всегда, а здесь тем более, применение математики имеет смысл только в единении с глубокой теорией конкретного явления. Об этом важно помнить, чтобы не сбиваться на простую игру в формулы, за которой не стояло бы никакого реального содержания. Но так или иначе, математика находит приложения почти во всех науках, от механики до политической экономии.

Напомним несколько примеров особенно блестящих применений математики в точных науках и технике.

Одна из самых далеких планет солнечной системы Нептун была открыта в 1846 г. на основании математических расчетов. Анализируя неправильности в движении планеты Уран, астрономы Адамс и Леверье пришли к выводу, что неправильности эти вызваны притяжением другой планеты. Леверье на основании законов механики и закона тяготения вычислил, где эта планета должна была находиться, и наблюдатель, которому он об этом сообщил, увидел ее в телескоп там, где указал Леверье. Это открытие было не только триумфом механики и астрономии, особенно системы Коперника, но также триумфом математического расчета.

Другой, не менее убедительный пример представляет открытие электромагнитных волн. Английский физик Максвелл, обобщая установленные опытами законы электромагнитных явлений, выразил эти законы в виде уравнений. Из уравнений он чисто математически вывел, что могут существовать электромагнитные волны и что они должны распространяться со скоростью света. Опираясь на это, он предложил электромагнитную теорию света, которая затем была всесторонне развита и обоснована. Но, кроме того, вывод Максвелла толкнул на поиски электромагнитных волн чисто электрического происхождения, например испускаемых при колебательном разряде. Такие волны были действительно открыты Герцем. А вскоре А. С. Попов нашел средства возбуждения, передачи и приема электромагнитных колебаний, вывел их в область широких применений и положил тем самым начало всей радиотехнике. В открытии радио, ставшего общим достоянием, сыграли большую роль также результаты чисто математического вывода.

Так от наблюдений,— каковы, например, наблюдения отклонений магнитной стрелки электрическим током,— наука идет к обобщению, к теории явлений, к формулировке законов и их математическому выражению. Из этих законов рождаются новые выводы, и, наконец, теория воплощается в практике, которая в свою очередь дает теории новые мощные импульсы к развитию.

Особенно замечательно, что даже самые абстрактные построения математики, возникшие внутри нее самой, уже без непосредственных толчков со стороны естествознания или техники, находят тем не менее плодотворные применения. Например, мнимые числа появились на свет в алгебре, и долгое время их реальный смысл оставался непонятным, на что показывает само их название. Однако после того, как в начале прошлого столетия им было дано геометрическое толкование (см. главу IV, § 2), мнимые числа вполне укрепились в математике, и возникла обширная теория функций комплексной переменной (т. е. переменной вида $x+y\sqrt{-1}$). Эта теория, так сказать, «мнимых» функций от «мнимых» переменных оказалась вовсе не мнимым, а очень реальным средством решения вопросов техники. Так, основная теорема Н. Е. Жуковского о подъемной силе крыла самолета доказывается как раз средствами этой теории. Та же теория оказывается полезной, например, при решении задач о просачивании воды под плотинами,— задач, значение которых очевидно в период строительства крупных гидроэлектростанций.

Другой, не менее блестящий пример представляет неэвклидова геометрия[1]. Она возникла на почве тысячелетних, тянувшихся со времен Эвклида попыток доказать аксиому о параллельных, т. е. из задачи, имевшей чисто математический интерес. Н. И. Лобачевский, создавший эту новую геометрию, сам осторожно называл ее «воображаемой», так как не мог указать ее реального значения, хотя и был уверен в том, что такое значение ее найдется. Выводы его геометрии казались большинству не то что «воображаемыми», но даже невообразимыми и нелепыми. Тем не менее идеи Лобачевского положили начало новому развитию геометрии, созданию теорий разных неэвклидовых пространств; потом эти идеи послужили одной из основ общей теории относительности, причем математическим аппаратом этой теории служит одна из форм неэвклидовой геометрии четырехмерного пространства. Так, казавшиеся по меньшей мере непонятными абстрактные построения математики оказались мощным средством развития одной из важнейших физических теорий. Точно так же в современной теории атомных явлений, в так называемой квантовой механике, существенно используются многие чрезвычайно абстрактные математические понятия и теории, как, например, понятие бесконечномерного пространства и др.

Нет нужды углубляться в перечисление примеров; мы достаточно подчеркнули, что математика имеет широчайшее применение в повседневной практике, в технике, в науке, причем в точных науках и больших проблемах техники находят также применения теории, выросшие внутри самой математики. Такова одна из характерных особенностей математики наряду с ее отвлеченностью, строгостью и убедительностью ее выводов.

2. Обратив внимание на все эти особенности математики, мы, конечно, не выяснили ее сущности, а указали, скорее, ее внешние признаки. Задача состоит в том, чтобы объяснить эти особенности. Для этого нужно ответить, по крайней мере, на следующие вопросы:

Что отражают абстрактные математические понятия? Каков, иными словами, реальный предмет математики?

---

[1] Здесь мы только указываем этот пример, не вдаваясь в объяснения, которые читатель найдет в главе XVII (том 3).

Почему отвлеченные математические выводы представляются столь убедительными, а первичные понятия столь очевидными? В чем, иными словами, основание метода математики?

Почему при всей своей отвлеченности математика находит широчайшее применение, а не оказывается праздной игрой в абстракции? Иными словами: откуда значение математики?

Наконец, какие силы двигают развитие математики, позволяя ей соединять абстрактность и широту применений? Иными словами: в чем содержание процесса развития математики?

Ответив на эти вопросы, мы получим общее предотавление о предмете математики, об основаниях ее метода, о ее значении и развитии, т. е. поймем ее сущность.

Идеалисты и метафизики не только путаются в решении этих коренных вопросов, но доходят до полного извращения математики, выворачивая ее в буквальном смысле наизнанку. Так, видя крайнюю отвлеченность и убедительность математических выводов, идеалисты воображают, что математика происходит из чистого мышления.

В действительности математика не дает никаких оснований для идеализма и метафизики; как раз наоборот: рассматриваемая объективно во всех ее связях и развитии, она дает еще одно блестящее подтверждение диалектического материализма и каждым своим шагом опровергает идеализм и метафизику. Мы убедимся в этом, когда попытаемся даже в самых общих чертах ответить на поставленные выше вопросы о сущности математики. Мы убедимся также, что ответ на эти вопросы уже заключается в положениях, установленных классиками марксизма как относительно математики, так и относительно природы науки и познания вообще. Для предварительного выяснения этих вопросов достаточно рассмотреть основания арифметики и элементарной геометрии. К ним мы и обратимся. Дальнейшее проникновение в математику, конечно, углубит и разовьет, но никак не отменит полученные при этом выводы.

§ 2. АРИФМЕТИКА

1. Понятие о числе (мы говорим пока только о целых положительных числах) — это понятие, такое для нас привычное, вырабатывалось очень медленно. Об этом можно судить хотя бы по тому, как считали народы, еще совсем недавно стоявшие на разных ступенях первобытно-общинного строя. У некоторых из них не было даже названий для чисел больше двух или трех, у других счет шел дальше, но так или иначе он сравнительно быстро кончался, и о большем числе они говорили просто «много» или «неисчислимо». Это показывает, что запас ясно различаемых чисел накапливался у людей постепенно.

Вначале люди не имели понятия о числе, хотя и могли по-своему судить о размерах той или иной совокупности вещей, встречавшейся в их практике. Надо думать, что число воспринималось ими непосредственно как неотъемлемое свойство совокупности предметов, которое, однако, еще ими явно не выделялось. Мы настолько привыкли к счету, что едва ли можем себе это представить, но понять это можно [1].

На более высокой ступени число уже указывается как свойство совокупности предметов, но еще не отделяется от нее как «отвлеченное число», как число вообще, не связанное с конкретными предметами. Это видно из таких названий чисел у некоторых народов, как «рука» для пяти, «весь человек» — для двадцати и т. п. Здесь пять понимается не отвлеченно, а просто как «столько же, сколько пальцев на руке», двадцать — как «столько же, сколько всех пальцев у человека» и т. п. Совершенно аналогично у некоторых народов не было, например, понятий «черный», «твердый», «круглый». Чтобы сказать, что предмет черный, они сравнивали его, допустим, с вороном, а чтобы сказать, что имеется пять предметов, они прямо сравнивали эти предметы с рукой. Бывало и так, что разные названия чисел употреблялись для разного рода предметов: одни числа для счета людей, другие для счета лодок и т. д. до десяти разных сортов чисел. Тут нет отвлеченных чисел, они являются как бы «именованными», относящимися только к определенному роду предметов. У других народов вообще нет отдельных названий для чисел, например нет слова «три», хотя они могут сказать: «три человека», «в трех местах» и т. п.

Аналогично этому мы легко говорим, что тот или иной предмет черный, но гораздо реже говорим о «черноте» самой по себе,— это понятие представляется более абстрактным[2].

Число предметов есть свойство некоторой их совокупности, число же, как таковое, иными словами «отвлеченное число», есть это свойство, отвлеченное от конкретных совокупностей и мыслимое уже само по себе, подобно «черноте», «твердости» и т. п. Как чернота есть общее свойство предметов цвета угля, так число «пять» есть общее свойство всех совокупностей, содержащих столько же предметов, сколько пальцев на руке.

---

[1] В самом деле, всякая совокупность предметов, будь то стадо овец или поленница дров, существует и непосредственно воспринимается во всей своей конкретности и сложности. Выделение в ней отдельных свойств и отношений есть результат известного анализа. Примитивное мышление еще не делает такого анализа, а берет объект только в целом. Подобно этому, например, человек, не занимавшийся музыкой, воспринимает музыкальное произведение, не выделяя в нем деталей мелодии, тональности и т. п., в то время как музыкант легко анализирует даже сложную симфонию.

[2] В образовании понятий о свойствах предметов, будь то цвет или численность совокупности, можно различить три ступени, которые, впрочем, нельзя слишком строго разграничивать. На первой ступени свойство определяется прямым сравнением предметов: такой же, как ворон; столько же, сколько на руке. На второй ступени появляется прилагательное: черный камень, аналогично — числительное: пять деревьев и т. п. На третьей ступени свойство отвлекается от предметов и может фигурировать «как таковое», как «чернота», как отвлеченное число «пять» и т. п.

При этом сама равночисленность устанавливается простым сравнением: беря предмет из совокупности, мы загибаем один палец и так пересчитываем их по пальцам. Вообще сопоставлением предметов двух совокупностей можно, вовсе не пользуясь числами, установить, одинаковое ли в них число предметов. Так, гости, рассаживаясь за столом и ничего не считая, легко поправляют хозяйку, если она забыла один прибор: один гость остался без прибора.

Таким образом, можно дать следующее определение числа: *каждое отдельное число, как «два», «пять» и т. п., есть свойство совокупностей предметов, общее для всех совокупностей, предметы которых можно сопоставить по одному, и различное у таких совокупностей, для которых такое сопоставление невозможно.*

Для того чтобы обнаружить и ясно выделить это общее свойство, т. е. для того чтобы образовать понятие о том или ином числе и дать ему название «шесть», «десять» и т. д., нужно было сравнить между собой немало совокупностей предметов. Люди считали на протяжении долгих поколений, миллионы раз повторяя одни и те же операции, и так на практике обнаруживали числа и отношения между ними.

2. Действия, операции над числами возникали, в свою очередь, как отражение реальных действий над конкретными предметами. Это заметно и в названиях чисел. Так, например, у некоторых индейцев число «двадцать шесть» произносится, как «на два десятка я кладу сверху шесть». Ясно, что здесь отражается конкретный способ пересчитывания предметов. Тем более ясно, что сложение чисел соответствует складыванию, соединению двух или нескольких совокупностей в одну. Так же легко видеть конкретный смысл вычитания, умножения и деления (умножение, в частности, в большой мере происходит, видимо, от счета равными совокупностями: по 2, по 3 и т. п.).

В процессе счета люди открывали и усваивали не только связи между отдельными числами, как, например, то, что два и три будет пять, но устанавливали постепенно и общие законы. На практике обнаруживалось, что сумма не зависит от порядка слагаемых или что результат счета данных предметов не зависит от того, в каком порядке этот счет производится. (Это последнее обстоятельство находит выражение в совпадении «порядковых» и «количественных» чисел: первый, второй и т. д. и один, два и т. д.) Таким образом, числа выступали не как отдельные и независимые, а в связи друг с другом.

Одни числа выражаются через другие даже в названиях и записи. Так, «двадцать» означает «два (раза) десять», по-французски 80 — «четыре-двадцать» (quatre-vingt), 90 — «четыре-двадцать-десять», а, например, римские цифры VIII, IX означают, что $8=5+3$, $9=10-1$.

В общем, возникали не просто отдельные числа, а *система* чисел с ее связями и законами.

Предмет арифметики составляет именно система чисел с ее связями и законами[1]. Отдельное отвлеченное число само по себе не имеет содержательных свойств, и о нем вообще мало что можно сказать. Если мы спросим себя, например, о свойствах числа 6, то заметим, что $6=5+1$, $6=3\cdot 2$, что 6 есть делитель 30 и т. п. Но здесь всюду число 6 связывается с другими числами, так что свойства данного числа состоят именно в его отношениях к другим числам[2]. Тем более ясно, что всякое арифметическое действие определяет связь, или, иными словами, отношение между числами.

Таким образом, арифметика имеет дело с отношениями между числами. Но отношения между числами являются отвлеченными образами реальных количественных отношений между совокупностями предметов, поэтому можно сказать, что *арифметика есть наука о реальных количественных отношениях, рассматриваемых, однако, отвлеченно, так сказать, в чистом виде.*

Арифметика, как мы видим, происходит не из чистого мышления, как стараются изобразить идеалисты, а отражает определенные свойства реальных вещей; она возникла в результате долгого практического опыта многих поколений.

3. Чем обширнее и сложнее становилась общественная практика, тем более широкие задачи она ставила. Нужно было не только отмечать количество предметов и обмениваться мыслями об их числе, что уже потребовало формирования понятия числа и названий чисел, но надо было учиться считать все бо́льшие совокупности (будь то животные в стаде, предметы при обмене, дни до намеченного срока и т. п.), фиксировать и передавать другим результаты счета, что как раз и требовало совершенствования названий, а затем и обозначений для чисел.

Введение обозначений для чисел, идущее, повидимому, от самого зарождения письменности, сыграло громадную роль в развитии арифметики. Кроме того, это был первый шаг к математическим знакам и формулам вообще. Следующий шаг, состоявший во введении знаков для арифметических действий и буквенного обозначения для неизвестного (x), был сделан гораздо позже.

Понятие числа, как всякое абстрактное понятие, не имеет одного непосредственного образа, его нельзя представить, а можно только мыслить. Но мысль оформляется в языке, поэтому без названия нет и поня-

---

[1] Слово «арифметика» происходит от греческого «искусство счета» («арифмос» — число и «техне» — искусство).

[2] Это понятно также из самых общих соображений. Любая абстракция, отделенная от конкретного основания,— как число отвлекается от конкретных совокупностей предметов,— «сама по себе» не имеет смысла, она живет только в связях с другими понятиями. Эти связи содержатся уже в любом высказывании, в самом неполном определении. Вне их она лишается содержания и значения, т. е. просто не существует. Содержание понятия отвлеченного числа лежит в законах, в связях системы чисел.

тия. Обозначение есть то же название, только не звуковое, а письменное, и оно воспроизводится мыслью в виде зрительного образа. Например, если я скажу «семь», что вы представляете? Вероятно, не семь каких-нибудь предметов, а прежде всего цифру «7»; она и служит материальной оболочкой для отвлеченного числа «семь». А такое число, как, например, 18 273, заметно труднее произнести, чем написать, и уж вовсе нельзя с полной точностью представить себе в образе совокупности предметов. Таким образом, обозначения помогли, хотя и не сразу, создать понятие о таких числах, которых уже нельзя было обнаружить в простом наблюдении и непосредственном пересчитывании. В этом была практическая необходимость: с появлением государства нужно было собирать подати, собирать и снабжать войско и т. п., что требовало операций с очень большими числами.

Итак, во-первых, роль обозначений для чисел состоит в том, что они дают простое воплощение понятия отвлеченного числа[1]. Такова роль математических обозначений вообще: они дают воплощение отвлеченных математических понятий. Так, $+$ означает сложение, x — неизвестное число, a — любое данное число и т. д. Во-вторых, обозначения чисел дают возможность особенно просто осуществлять действия над ними. Каждый знает, насколько легче «подсчитать на бумаге», чем «в уме». Такое же значение имеют математические знаки и формулы вообще: они позволяют заменять часть рассуждений выкладками, т. е. чем-то почти механическим. К тому же, если выкладка написана, она имеет уже определенную достоверность. Тут все видно, все можно проверить, все определяется точными правилами. Для примера можно вспомнить сложение «столбиком» или любое алгебраическое преобразование, как, например, «перенесение в другую часть равенства с изменением знака».

Из сказанного ясно, что без подходящих обозначений для чисел арифметика не могла бы продвинуться далеко вперед. Тем более современная математика была бы просто невозможна без специальных знаков и формул.

Само собой понятно, что люди далеко не сразу смогли выработать современный, столь удобный, способ записи чисел. С древних времен у разных народов с начатками культуры появлялись разные числовые обозначения, мало похожие на наши современные не только по начертанию знаков, но и по принципам; не всюду, например, пользовались именно десятичной системой (так, у древних вавилонян была смешанная десятичная и шестидесятиричная система). На прилагаемой таблице показаны для примера

---

[1] Стоит заметить, что понятие о числах, которое вырабатывалось, как мы видели, с таким трудом в течение очень долгого времени, усваивается теперь ребенком сравнительно легко. Почему? Во-первых, конечно, потому, что ребенок слышит и видит, как взрослые постоянно пользуются числами, и они даже учат его этому. А во-вторых, потому,— и именно на это мы хотим обратить внимание,— что ребенок имеет готовые слова и обозначения для чисел. Он сначала выучивает эти внешние образы числа, а потом уже овладевает его смыслом.

	Славянское		Китайское			Греческое	Арабское	Грузинское	Египетское		Римское	Цифры племени Майя
	Кириллицей	Глаголицей	Старое	Коммерческое	Научное				Иероглифы	Иератическое		
0				○	○							𝋠
1	а҃	Ⰰ	一	〡	𝍠	α̅	ا	ა	𓏺	[illegible]	I	𝋡
2	в҃	Ⰱ	二	〢	𝍡	β̅	ب	ბ	𓏻	[illegible]	II	𝋢
3	г҃	Ⰲ	三	〣	𝍢	γ̅	ج	გ	𓏼	[illegible]	III	𝋣
4	д҃	Ⰳ	四	〤	𝍣	δ̅	د	დ	𓏽	[illegible]	IV	𝋤
5	є҃	Ⰴ	五	〥	𝍤	ε̅	ه	ე	𓏾	[illegible]	V	𝋥
6	ѕ҃	Ⰵ	六	〦	𝍥	ϛ̅	و	ვ	𓏿	[illegible]	VI	𝋦
7	з҃	Ⰶ	七	〧	𝍦	ζ̅	ز	ზ	𓐀	[illegible]	VII	𝋧
8	и҃	Ⰷ	八	〨	𝍧	η̅	ح	ჱ	𓐁	[illegible]	VIII	𝋨
9	ѳ҃	Ⰸ	九	〩	𝍨	θ̅	ط	თ	𓐂	[illegible]	IX	𝋩
10	і҃	Ⰹ	十	十	𝍠○	ι̅	ي	ი	𓎆	Λ	X	𝋪
20	к҃	Ⰻ	二十	〢十	𝍡○	κ̅	ك	კ	𓎇	[illegible]	XX	
30	л҃	Ⰼ	三十	〣十	𝍢○	λ̅	ل	ლ	𓎈	[illegible]	XXX	
100	р҃	Ⱃ	百	[illegible]	𝍠○○	ρ̅	ق	რ	𓍢	[illegible]	C	
1000	҂а	Ⱍ	千	千	𝍠○○○	͵α	غ	ჩ	𓆼	[illegible]	M	

Обозначения чисел у разных народов.

Заимствовано из статьи И. Г. Башмаковой и А. П. Юшкевича «Происхождение систем счисления» («Энциклопедия элементарной математики», т. I, М., 1951) некоторые из обозначений чисел у разных народов. В частности, мы видим, что древние греки, а потом и русские пользовались алфавитными обозначениями. Наши современные «арабские» цифры и вообще способ записи чисел происходят из Индии, откуда они были занесены арабами в X в. в Европу, где окончательно укоренились в течение нескольких столетий.

Первая особенность нашей системы состоит в том, что она десятичная. Но особенность эта не очень существенная, потому что можно было бы пользоваться с успехом, скажем, двенадцатиричной системой, введя особые обозначения для десяти и одиннадцати.

Главная особенность нашей системы обозначений состоит в том, что она — «позиционная», т. е. в ней одна и та же цифра имеет разное значение в зависимости от занимаемого места. Так, например, в обозначении 372 цифра 3 обозначает число сотен, а 7 — число десятков. Такой способ записи не только краток и прост, но крайне облегчает вычисления. Римские обозначения куда менее удобны: то же число 372 запишется по-римски так: CCCLXXII, а умножать большие числа, записанные по-римски, совсем неудобно.

Позиционная запись чисел требует, чтобы как-то отмечался пропущенный разряд, так как если его не отмечать, то мы будем путать, например, триста один и тридцать один. На месте пропущенного разряда ставится нуль; так мы различаем 301 и 31. В зачаточном виде нуль появляется уже в поздних вавилонских клинописях. Систематическое введение нуля было достижением индийцев[1]: оно-то и позволило им довести до конца позиционную систему записи чисел, которой мы сейчас пользуемся.

Но этого мало: нуль стал тоже числом, войдя в систему чисел. Сам по себе нуль есть ничто — на санскритском (древнеиндийском) языке он так и называется: «пустой» (çūnga), но в связи с другими числами нуль приобретает содержание, приобретает известные свойства — хотя бы то, что любое число плюс нуль есть то же число, а умноженное на нуль есть нуль.

4. Вернемся к арифметике древних. Старейшие дошедшие до нас математические тексты из Вавилона и Египта восходят ко второму тысячелетию до н. э. Эти и более поздние тексты содержат разнообразные арифметические задачи с решениями, и притом даже такие, которые относятся теперь к алгебре, как решения некоторых квадратных и даже кубических уравнений или прогрессии (все это, конечно, на конкретных задачах и численных примерах). Из Вавилона дошли до нас также таблицы квадратов, кубов и обратных чисел. Есть предположение, что там уже складывались математические интересы, не связанные непосредственно с отдельными практическими задачами.

[1] Первая индийская рукопись, где находят нуль, относится к концу IX в.; в ней есть запись числа 270 точно в наших обозначениях. Однако нуль, вероятно, был введен в Индии еще раньше — в VI в.

Во всяком случае в древнем Вавилоне и Египте арифметика была хорошо развита. Но она не была еще математической теорией чисел, а набором правил счета и решения различных задач. Так, впрочем, преподают арифметику в начальной школе по настоящее время и так же понимают ее все, кто не занимался специально математикой. Это вполне законно, но все же в таком виде арифметика не есть еще математическая теория: в ней нет общих теорем о числах.

Переход к теоретической арифметике происходил постепенно.

Обозначения, как мы говорили, дают возможность оперировать с такими большими числами, которые уже нельзя наглядно представить в виде совокупности предметов и до которых нельзя дойти, считая подряд от единицы. Если у диких племен числа обрываются на 3, 10, 100 и т. п., а дальше следует неопределенное «много», то обозначения дали возможность в Китае, Вавилоне, Египте идти за десятки тысяч и даже за миллионы. Тут уже намечалась возможность неограниченного продолжения числового ряда. Но ясно осознана она была не сразу, когда точно, мы не знаем. Еще Архимед (287—212 гг. до н. э.) в своем знаменитом сочинении «Об исчислении песка» указывал способ назвать число, большее числа песчинок, которое могло бы уместиться в «шаре неподвижных звезд». Возможность назвать и записать такое число, стало быть, еще требовала в то время подробного разъяснения.

Греки к III веку до н. э. уже ясно осознали две важные идеи: во-первых, что ряд чисел можно неограниченно продолжать, и, во-вторых, что можно не только оперировать с любыми данными числами, но и рассуждать о числах вообще, формулируя и доказывая общие теоремы о числах. Это было обобщением огромного предшествующего опыта в оперировании с конкретными числами. Именно из этого опыта выявились общие законы и приемы *общих* рассуждений о числах. Произошел переход на более высокую ступень абстракции: от отдельных данных (хотя и отвлеченных) чисел к числу вообще, к *любому возможному числу*.

От простого процесса пересчитывания предметов по одному мы переходим к представлению о неограниченном процессе образования чисел путем прибавления единицы к ранее построенному числу. Ряд чисел мыслится уже неограниченно продолжаемым, и с ним в математику вступает бесконечность. Конечно, мы фактически не можем путем прибавления единиц зайти сколь угодно далеко в ряду чисел: кто сможет досчитать до миллиона миллионов, если даже сто лет содержат почти в 40 раз меньше секунд? Но не в том дело. Процесс накопления единиц, процесс образования сколь угодно больших совокупностей предметов принципиально не ограничен, и, стало быть, есть *потенциальная возможность* неограниченного продолжения числового ряда. Практическая ограниченность счета тут не при чем, от нее отвлекаются. Общие теоремы о числах касаются уже этого неограниченно продолжаемого ряда чисел.

Общие теоремы о каком-либо свойстве любого числа уже содержат в скрытом виде бесконечно много утверждений о свойствах отдельных чисел и качественно богаче каких-либо частных утверждений, которые можно было бы проверить для отдельных чисел. Поэтому общие теоремы необходимо должны доказываться путем общих рассуждений, исходящих из самого закона образования ряда чисел. Здесь открывается глубокая особенность математики: математика имеет своим предметом не только данные количественные отношения, но вообще возможные количественные отношения и, стало быть, бесконечность.

В знаменитых «Началах» Эвклида, написанных в III в. до н. э., есть уже общие теоремы о целых числах, в частности теорема о том, что существуют сколь угодно большие простые числа[1].

Так арифметика превращается в теорию чисел. Она уже отвлекается от конкретных частных задач и вращается в области отвлеченных понятий и рассуждений. Она становится частью «чистой» математики. Вернее, это и был момент рождения самой чистой математики со всеми ее особенностями, о которых шла речь в п. 1. Нужно, правда, заметить, что чистая математика рождалась одновременно из арифметики и геометрии. Кроме того, в общих правилах арифметики имелись уже зачатки алгебры, которая отделилась от арифметики позже. Но об этом мы будем говорить ниже.

Теперь же остается только подвести итоги всех наших выводов, потому что мы хотя и очень бегло, но все же проследили процесс возникновения теоретической арифметики от самого зарождения понятия о числе.

5. Поскольку рождение теоретической арифметики было частью рождения математики, мы, естественно, можем ожидать, что наши выводы об арифметике осветят общие вопросы, касающиеся математики вообще. Вспомним эти вопросы, применяя их к арифметике.

1° Как возникают и что отражают в действительности отвлеченные понятия арифметики?

На этот вопрос отвечает все, что было рассказано о зарождении арифметики. Ее понятия отражают количественные отношения совокупностей предметов. Возникали эти понятия путем абстракции, на основе анализа и обобщения громадного практического опыта. Они возникали при этом постепенно; сначала числа, связанные с конкретными предметами, потом отвлеченные числа и, наконец, понятие о числе вообще, о *любом возможном числе*. Каждая из этих ступеней подготовлялась накоплением опыта с применением предыдущих понятий. (Таков, кстати, один из основных законов образования математических понятий: они рождаются путем последовательного ряда абстракций и обобщений, опирающихся на накопленный опыт применения предшествующих отвлеченных понятий).

---

[1] Напоминаем, что простыми называются отличные от единицы положительные целые числа, которые делятся без остатка только на самоё себя и на единицу.

История возникновения понятий арифметики доказывает всю ошибочность идеалистических взглядов о том, будто эти понятия происходят из «чистого мышления», из «первоинтуиции», из «созерцания в априорных формах» и мало ли еще из чего.

2° Почему выводы арифметики представляются такими убедительными и непреложными?

История отвечает нам и на этот вопрос. Мы видим, что сами выводы арифметики вырабатывались медленно и постепенно; они отражают опыт, накапливавшийся в течение необозримо долгих поколений и таким путем закреплявшийся в сознании людей. Они закреплялись в языке: в названиях чисел, в обозначениях, в постоянном повторении одинаковых операций с числами, в постоянном их практическом применении. Так они приобрели ясность и убедительность. Самые приемы логических рассуждений имеют то же происхождение. При этом существенной является не только сама повторяемость, но и та устойчивость и четкость, которыми объективно обладают отношения действительности, отраженные в основных понятиях арифметики и в правилах логического вывода.

В этом корень убедительности арифметики; ее выводы логически вытекают из ее основных понятий, а то и другое — приемы логики и понятия арифметики — вырабатывались и закреплялись в сознании на основе тысячелетней практики, на основе объективных закономерностей окружающего нас мира.

3° Почему арифметика при всей отвлеченности ее понятий имеет такие широкие приложения?

Ответ прост. Понятия и выводы арифметики, обобщая огромный опыт, выражают в отвлеченной форме такие отношения действительности, которые встречаются постоянно и повсюду. Считать можно и вещи в комнате, и звезды, и людей, и атомы... Арифметика берет некоторые из общих свойств, отвлекаясь от всего частного и конкретного. И именно в силу того, что она берет только это общее, ее выводы приложимы в такой массе случаев. Стало быть, возможность широких приложений обеспечивается именно отвлеченностью арифметики. (При этом важно, что отвлеченность эта не пустая, а извлечена из долгого практического опыта.) То же верно в отношении всей математики, в отношении любого отвлеченного понятия или теории. Возможности приложения теории зависят от того, сколь широкий исходный материал в ней обобщен.

Одновременно всякое отвлеченное понятие, в частности понятие о числе, ограничено в своем значении вследствие той же самой своей отвлеченности. Во-первых, в применении к любому конкретному предмету оно отражает только одну его сторону и поэтому дает о нем очень неполное представление. Как часто, например, бывает, что одни численные данные еще очень мало говорят по существу дела. Во-вторых, отвлеченные понятия нельзя применять всюду без каких бы то ни было условий, нельзя применять арифметику к конкретным задачам, не убедившись в том, что ее применение имеет здесь смысл. Если мы, например, говорим о сложении, соединяя предметы только мысленно, то, конечно, с самими предметами ничего не происходит. Но если мы применяем сложение к фактическому соединению предметов, если мы фактически складываем предметы, например, сваливая их в кучу или расставляя на столе, то здесь происходит не просто отвлеченное сложение, а реальный процесс. Этот процесс не только не исчерпывается арифметическим сложением, но может сделать его вообще неприменимым. Например, сваливаемые в кучу предметы могут ломаться; звери, посаженные вместе, могут растерзать один другого; «складываемые» вещества могут вступить в химическую реакцию: литр воды и литр спирта дадут при слиянии не 2, а 1,9 литра смеси вследствие взаимного растворения этих жидкостей, и т. п.

Нужны ли другие примеры? Их можно привести сколько угодно.

Короче, истина конкретна; и помнить это особенно важно в отношении математики, как раз из-за ее отвлеченности.

4° Наконец, последний вопрос, который мы ставили, касался движущих сил развития математики.

Для арифметики ответ на этот вопрос также ясен из истории ее возникновения. Мы видели, что люди на практике овладевали счетом и вырабатывали понятие о числе; потом практика потребовала обозначений для чисел, поставила более трудные задачи. Словом, движущей силой развития арифметики служила общественная практика. При этом она выступает в постоянном взаимодействии с обобщающим ее опыт отвлеченным мышлением. Возникающие на основе практики отвлеченные понятия делаются важным ее орудием и совершенствуются в своем применении. Отвлечение от несущественного помогает при этом вскрывать суть дела и обеспечивает общее решение там, где определяющую роль играют как раз выделенные и сохраненные при отвлечении общие свойства и связи,— таковы количественные связи в случае арифметики.

Кроме того, мышление зачастую уходит дальше того, что непосредственно требует поставленная практикой задача. Так, понятие о больших числах, как миллион или миллиард, возникло на базе счета, но раньше, чем явилась практическая потребность пользоваться такими числами. Таких примеров не мало в истории науки; достаточно вспомнить мнимые числа, о которых мы уже упоминали. Все это лишь частный случай общего всему познанию взаимодействия практики и абстрактного мышления, практики и теории.

§ 3. ГЕОМЕТРИЯ

1. История зарождения геометрии по существу сходна с историей зарождения арифметики. Первые геометрические понятия и сведения также восходят к временам доисторическим и также возникли в процессе практической деятельности.

Из самой природы заимствовал человек геометрические формы. Круг и серп луны, гладь озера, прямизна луча или стройного дерева существовали задолго до человека и предстояли перед ним постоянно. Конечно, достаточно прямые линии, тем более треугольники и квадраты, наш глаз встретит в самой природе редко. Ясно, что человек, вырабатывал представление об этих фигурах прежде всего потому, что активно воспринимал природу и, следуя своим практическим потребностям, изготовлял предметы все более и более правильной формы. Люди строили свои жилища, обтесывали камни, огораживали участки земли, натягивали тетивы на свои луки, лепили глиняную посуду, совершенствовали ее и соответственно создавали понятие, что сосуд получается *круглый*, что натянутая тетива — *прямая*. Короче, форму сначала придавали материалу, а потом уже осознавали ее как то, что придается материалу и что может рассматриваться само по себе в отвлечении от материала. Осознавая форму тел, человек мог совершенствовать свои изделия и еще отчетливее выделять само понятие формы. Так, практическая деятельность служила основой для выработки отвлеченных понятий геометрии. Нужно было сделать тысячи предметов с прямыми краями, натянуть тысячи нитей, провести на земле массу прямых линий, чтобы получить ясное представление о прямой линии вообще, как о том общем, что есть во всех этих частных случаях. Теперь мы окружены предметами с прямыми краями, сделанными людьми, сами учимся проводить прямые, и только поэтому у нас с детства складывается ясное представление о прямой.

Точно так же понятие о геометрических величинах — о длине, площади и объеме — возникло из практической деятельности. Люди измеряли длины, определяли расстояния, оценивали на глаз площади и объемы для своих практических целей. Постепенно здесь были обнаружены простейшие общие законы, первые геометрические зависимости, например: площадь прямоугольника равна произведению его сторон. Земледельцу полезно было знать такую зависимость, чтобы оценивать площадь посева, а следовательно, и предполагаемый урожай.

Так из практической деятельности и жизненных задач зарождалась геометрия. Вот что писал о ней в IV в. до н. э. древнегреческий ученый Эвдем Родосский: «Геометрия была открыта египтянами и возникла при измерении земли. Это измерение было им необходимо вследствие разлития реки Нила, постоянно смывавшего границы[1]. Нет ничего удивительного в том, что эта наука, как и другие, возникла из потребностей человека. Всякое возникающее знание из несовершенного состояния переходит в совершенное. Зарождаясь путем чувственного восприятия, оно постепенно становится предметом нашего рассмотрения и, наконец, делается достоянием разума».

---

[1] Речь идет о границах земельных участков. Заметим, кстати, что *геометрия* и означает в переводе землемерие (по-древнегречески «ге»—земля, «метрео»— мерю).

Конечно, измерение земли не было единственной задачей, побудившей древних к созданию геометрии. О характере задач и о том, как их решали древние египтяне и вавилоняне, можно судить по дошедшим до нас отрывкам текстов. Один из самых древних, дошедших до нас, египетских текстов восходит к временам более 1700 лет до н. э.— это руководство «писцам» (царским чиновникам), написанное неким Ахмесом. Здесь собран ряд задач на вычисление вместимости сосудов и амбаров, площадей земельных участков, размеров земляных работ и др.

Египтяне и вавилоняне умели определять простейшие площади и объемы, знали с хорошей точностью отношение окружности к диаметру и, может быть, даже могли вычислять поверхность шара, словом, они обладали уже немалым запасом геометрических знаний. Однако, насколько можно судить, у них не было еще геометрии как теоретической науки с ее теоремами и доказательствами. Как и арифметика того времени, геометрия была в основном набором правил, выведенных из опыта. Более того, геометрия вообще не была отделена от арифметики. Геометрические задачи были одновременно арифметическими задачами на вычисление.

В VII в. до н. э. геометрия проникла из Египта в Грецию, где ее развивали дальше великие философы-материалисты Фалес, Демокрит и другие. Значительный вклад в геометрию сделали также последователи Пифагора — основателя идеалистической религиозно-философской школы.

Развитие геометрии шло в направлении накопления новых фактов и уяснения их связей друг с другом. Эти связи превращались постепенно в логические выводы одних положений геометрии из других. Таким путем, во-первых, вырабатывалось самое понятие о геометрической теореме и ее доказательстве, а во-вторых, выяснялись те основные положения, из которых другие уже могут быть выведены, т. е. выяснялись аксиомы геометрии.

Так постепенно геометрия превращалась в математическую теорию.

Известно, что ее систематические изложения появились в Греции уже в V в. до н. э., но они не дошли до нас, очевидно, потому, что всех их вытеснили «Начала» Эвклида (III в. до н. э.). В этом произведении геометрия была представлена в виде такой стройной системы, что ничего принципиально нового к ее основам не смогли добавить до Н. И. Лобачевского, т. е. в течение более двух тысяч лет, а известный школьный учебник Киселева, как и многие другие учебники во всем мире, в старых изданиях представлял собой не что иное, как популярную переработку Эвклида. Едва ли много найдется в мире таких долговечных книг, как «Начала» Эвклида,— это совершенное творение греческого гения. Конечно, математика ушла вперед и наше понимание оснований геометрии стало гораздо глубже, и все же «Начала» Эвклида были и во многом остаются образцом книги по чистой математике. В них, подводя итог предыдущего развития, Эвклид представил современную ему математику как самостоятельную теоретическую науку, т. е. так, как в конце концов понимают ее и теперь.

2. История зарождения геометрии дает основание для тех же выводов, что история зарождения арифметики. Мы видим, что геометрия возникла из практики и что ее превращению в математическую теорию предшествовал очень долгий период.

Геометрия оперирует с «геометрическими телами» и фигурами, изучает их отношения величины и взаимного расположения. Но геометрическое тело есть не что иное, как реальное тело, рассматриваемое только с точки зрения его пространственной формы[1], в отвлечении от прочих свойств, будь то плотность, цвет, вес и т.д. (Геометрическая фигура есть еще более общее понятие, в ней можно отвлекаться и от пространственного протяжения; так, поверхность имеет только два измерения, линия — одно, а точка вовсе не имеет измерений. Точка есть отвлеченное понятие о конце линии, о месте, уточненном до предела, так что в нем уже нет частей. Так, кстати, определял все эти понятия и Эвклид.)

Таким образом, геометрия имеет своим предметом пространственные формы и отношения реальных тел, отвлеченные от всех прочих свойств, иными словами, взятые «в чистом виде». Именно этот уровень отвлеченности отличает геометрию от других наук, изучающих также пространственные формы и отношения тел. В астрономии, например, изучают взаимное расположение тел, но именно небесных тел, в геодезии — форму Земли, в кристаллографии — формы кристаллов и т. д. Во всех этих случаях изучают форму и расположение конкретных тел в связи и в зависимости от их других свойств.

Отвлечение влечет за собой умозрительный метод геометрии; с прямыми без всякой ширины, с «чистыми формами» уже нельзя ставить опыты. Остается только рассуждать, получая одни выводы из других. Поэтому геометрическая теорема должна доказываться рассуждением, иначе она не будет принадлежать геометрии, не будет относиться именно к «чистым формам».

Очевидность самих исходных понятий геометрии, приемы рассуждения, убедительность её выводов имеют то же происхождение, что и в арифметике. Свойства геометрических понятий, как и сами понятия, абстрагированы человеком из окружающей природы. Люди много раз проводили прямые линии, прежде чем смогли осознать как аксиому, что через всякие две точки можно провести прямую; миллиарды раз перемещали и прикладывали друг к другу разные предметы, прежде чем смогли обобщить это в представлении о наложении геометрических фигур и, тем более, применить это представление для доказательства теорем (как это делается в известных теоремах о равенстве треугольников).

Наконец, общность геометрии. Объем шара равен $^4/_3\pi R^3$ независимо от того, идет ли речь о шарообразном сосуде, о стальном шаре, о звезде, о капле и т. д. Геометрия смогла выделить *общее* всем телам, потому что

[1] В форму мы включаем также и размеры.

всякое реальное тело имеет более или менее определенные форму, размеры, положение относительно других тел. Не мудрено поэтому, что она применяется почти так же широко, как арифметика. Рабочий, измеряющий размеры детали или читающий чертеж, артиллерист, определяющий расстояние до цели, колхозник, измеряющий площадь посева, строитель, оценивающий объем земляных работ,— все они пользуются начатками геометрии. Штурман, астроном, геодезист, инженер, физик нуждаются в очень тонких ее выводах.

Яркий пример отвлеченно-геометрического решения важной задачи естествознания представляют исследования знаменитого кристаллографа и геометра Е. С. Федорова. Задача найти все возможные виды симметрии кристаллов, которую он перед собой поставил, является одной из основных в теоретической кристаллографии. Для решения этой задачи Федоров отвлекся от всех физических свойств кристалла, рассматривая его только как правильную систему геометрических тел (вместо системы конкретных атомов). Таким образом, речь идет о нахождении всех видов симметрии, какие только могут быть у системы геометрических тел. Этот чисто геометрический вопрос Федоров решил до конца и нашел все виды симметрии: их оказалось 230. Вместе с тем решение задачи о возможных видах симметрии явилось крупным вкладом в геометрию и послужило началом многих геометрических исследований.

На этом примере, как на всей истории геометрии, мы видим главную движущую силу ее развития. Это — взаимодействие практики и отвлеченного мышления. Возникшая из наблюдения кристаллов задача о их симметрии ставится отвлеченно и порождает новую математическую теорию — теорию правильных систем, или так называемых федоровских групп[1]. Потом сама эта теория не только находит блестящее подтверждение в наблюдении кристаллов, но служит общим руководством в развитии кристаллографии и порождает новые, как экспериментальные, так и чисто математические, исследования.

§ 4. АРИФМЕТИКА И ГЕОМЕТРИЯ

1. До сих пор мы рассматривали арифметику и геометрию отдельно друг от друга. Их взаимная связь, а стало быть, вообще связь математических теорий, ускользнула от нашего внимания. А между тем эта связь имеет исключительно большое значение. Взаимное проникновение теорий движет математику вперед, раскрывает богатство отраженных этими теориями связей действительности.

Арифметика и геометрия не только применяются одна к другой, но служат при этом истоками дальнейших общих идей, методов и теорий. В конечном счете арифметика и геометрия — это два корня, из которых росла математика. Их взаимодействие восходит к тем временам, когда сами они только зарождались. Уже простое измерение длины есть соединение геометрии и арифметики. Измеряя длину предмета, мы *откладываем* на нем некоторую единицу длины и *считаем*, сколько раз это можно сделать; первая операция (откладывание) — геометрическая, вторая (счет) — арифметическая. Отсчитывая шаги на дороге, каждый уже соединяет эти две операции.

---

[1] См. главу XX (том 3).

Вообще измерение любой величины соединяет счет с какой-либо специфической для этого сорта величин операцией. Достаточно вспомнить измерение жидкости мерным сосудом или измерение интервала времени по счету ударов маятника.

Однако при измерении обнаруживается, что, вообще говоря, выбранная единица не укладывается в измеряемой величине целое число раз и что простым счетом единиц при измерении обойтись нельзя. Тогда нужно единицу делить, чтобы выразить величину точнее посредством долей единицы, т. е. уже не целыми числами, а дробями. Дроби так фактически и возникли, как показывает анализ относящихся сюда исторических и других данных. Они возникли из деления и сравнения непрерывных величин, т. е. из измерения. Первые величины, которые люди измеряли, были геометрические величины: длины, площади посевов, объемы жидкостей и сыпучих тел. Стало быть, уже в возникновении дробей мы видим взаимодействие арифметики и геометрии. Это взаимодействие ведет к появлению важного нового понятия дроби, к расширению понятия числа от чисел целых к числам дробным (или, как говорят математики, к числам рациональным, выражающимся отношением целых чисел). Дроби не возникли и не могли возникнуть из деления целых чисел, потому что целыми числами считают целые предметы. Три человека, три стрелы и т. д.— все это имеет смысл, но две трети человека и даже треть стрелы представляет собой бессмыслицу; даже тремя отдельными третями стрелы не убить оленя,— для этого нужна *целая* стрела.

2. В развитии понятия числа, связанном с взаимодействием арифметики и геометрии, появление дробей было только первым этапом. Следующим было открытие несоизмеримых отрезков. Напомним, что отрезки называются несоизмеримыми, если не существует отрезка, который в обоих укладывается целое число раз, или, иными словами, если их отношение не выражается обычной дробью, т. е. отношением целых чисел.

Сначала люди просто не очень задумывались над тем, можно ли, скажем, всякую длину выразить дробью. Если при делении или измерении доходили до слишком мелких долей, их просто отбрасывали: бесконечное уточнение измерения практически не имело смысла. Демокрит выдвинул даже представление, что геометрические фигуры состоят из своего рода атомов. По Демокриту, отрезки—это ряды атомов, и потому отношение отрезков есть просто отношение чисел атомов в них, т. е. заведомо выражается дробью. Это представление, которое, на наш взгляд, может показаться довольно странным, оказалось очень плодотворным при определении площадей и объемов. Площадь вычислялась как сумма рядов, составленных из атомов, а объем — как сумма атомных слоев. Таким путем Демокрит нашел, например, объем конуса. Читатель, имеющий понятие об интегральном исчислении, легко заметит, что в этом приеме уже заключен прообраз определения площадей и объемов методами интегрального исчисления. (Кроме того, обращаясь мысленно к временам Демокрита, нужно постараться освободиться от ставших теперь привычными представлений, закрепленных развитием математики. Во время Демокрита геометрические фигуры еще не отрывались от реальных в той мере, в какой это делают теперь. Поэтому, мысля вещественные тела состоящими из атомов, Демокрит, естественно, должен был считать, что и геометрические фигуры состоят из атомов.)

Однако представление о том, что отрезки состоят из атомов, вступило в противоречие с теоремой Пифагора, так как из нее следует, что существуют несоизмеримые отрезки; например, диагональ квадрата не соизмерима с его стороной, т. е. отношение их не выражается отношением целых чисел.

Докажем, что сторона и диагональ квадрата действительно несоизмеримы. Если a — сторона, а b — диагональ квадрата, то по теореме Пифагора $b^2 = a^2 + a^2 = 2a^2$ и, стало быть,

$$\left(\frac{b}{a}\right)^2 = 2.$$

Между тем нет такой дроби, квадрат которой равнялся бы 2. В самом деле, допустим противное, и пусть p и q — целые числа, для которых

$$\left(\frac{p}{q}\right)^2 = 2,$$

причем мы можем, конечно, считать, что p и q уже не имеют общего делителя: иначе дробь можно было бы сократить. Но если $\left(\frac{p}{q}\right)^2 = 2$, то $p^2 = 2q^2$, и потому p^2 делится на 2. В таком случае p^2 делится на 4, ибо это есть квадрат целого числа. Поэтому $p^2 = 4q_1$, т. е. $2q^2 = 4q_1$, и $q^2 = 2q_1$. Отсюда следует, что q тоже должно делиться на 2. Это, однако, противоречит тому условию, что p и q не имеют общего делителя. Полученное противоречие доказывает, что отношение $\frac{b}{a}$ не может выражаться рациональным числом. Диагональ и сторона квадрата оказываются несоизмеримыми.

Это открытие произвело на греческих ученых громадное впечатление. Теперь, когда мы уже привыкли к иррациональным числам и легко действуем с квадратными и иными корнями, существование несоизмеримых отрезков нас нимало не смущает. Но в V в. до н. э. для греческих ученых открытие таких отрезков выглядело совершенно иначе. Ведь понятия иррационального числа у них не было, символа вроде $\sqrt{2}$ они не писали, и потому для них полученный результат означал, что отношение диагонали и стороны квадрата вообще не выражается никаким числом.

В существовании несоизмеримых отрезков грекам открывалась некая тайна, заключенная в непрерывности,— одно из выражений диалектического противоречия, заложенного в непрерывности и движении. Обсуждением этого противоречия занимались знаменитые греческие философы, среди которых особенно известен своими парадоксами Зенон Элейский.

Греки создали теорию отношений отрезков (и величин вообще), учитывающую существование несоизмеримых отрезков[1]; она изложена в «Началах» Эвклида, и в упрощенном виде ее излагают и теперь в школьном курсе геометрии. Но осознать, что отношение одного отрезка к другому, принятому за единицу, т. е. попросту длину отрезка, можно рассматривать тоже как число, обобщая этим самое понятие о числе,— до этой идеи греки так и не смогли подняться: понятие иррационального числа у них так и не возникло[2]. Это было сделано в более позднюю эпоху математиками Востока; а общее, математически строгое определение действительного числа, не опирающееся непосредственно на геометрию, было дано только совсем недавно: в 70-х годах прошлого столетия[3]. Такой громадный промежуток времени, протекший с тех пор, как было создана теория отношений, показывает, с каким трудом возникают и точно формулируются абстрактные понятия.

3. Характеризуя понятие о действительном числе, Ньютон в своей «Всеобщей арифметике», писал: «Под числом мы понимаем не столько собрание единиц, сколько отвлеченное отношение какого-либо количества к другому, принятому за единицу». Это число (отношение) может быть целым, рациональным или иррациональным, если данная величина не соизмерима с единицей.

Действительное, или, что то же самое, вещественное, число по своему исходному смыслу есть, следовательно, не что иное, как отношение одной величины к другой, принятой за единицу; в частном случае — это отношение отрезков, но может быть отношением площадей, весов и т. п.

Стало быть, действительное число есть отношение величин вообще, рассматриваемое в отвлечении от их конкретной природы.

Подобно тому, как *отвлеченные* целые числа становятся предметом математики не каждое в отдельности, а лишь в связи друг с другом, в си-

---

[1] Создание этой теории приписывается греческому ученому Эвдоксу, жившему в IV в. до н. э.

[2] Вследствие того, что учение об измерении величин не подчинялось арифметике, а входило в геометрию, геометрия у греков поглощала математику. Такие вопросы, как, например, решение квадратных уравнений, которые мы теперь трактуем алгебраически, они ставили и решали геометрически. «Начала» Эвклида содержат немало таких вопросов и, повидимому, представляли для современников сводку основ не одной геометрии в нашем смысле, но математики вообще. Это господство геометрии продолжалось до того, как Декарт, напротив, подчинил ее алгебре. Следы этого господства сохранились, например, в виде названий «квадрат» и «куб» для второй и третьей степени: «a куб» — это куб со стороной a.

[3] Речь идет не об описательном определении, а об определении, которое служит непосредственной основой для доказательств при изучении свойств действительных чисел. Естественно, что такие определения возникли в более позднюю эпоху, когда развитие математики, и главным образом анализа бесконечно малых, потребовало соответствующего определения действительного числа «переменной x». Это определение в разных формах было дано в 70-х годах прошлого века немецкими математиками Вейерштрассом, Дедекиндом и Кантором.

стеме целых чисел, так и *отвлеченные* действительные числа имеют содержание и оказываются предметом математики лишь в связи друг с другом, т. е. в системе действительных чисел.

В теории действительных чисел, как и в арифметике, прежде всего определяются действия над числами: сложение, вычитание, умножение, деление, а также отношение между числами, выражаемое словами «больше» и «меньше». Эти действия и отношения отражают реальные связи различных величин, как, например, сложение отражает сложение отрезков. Начало действий с отвлеченными действительными числами было положено в средние века математиками Востока. Позже было постепенно выявлено важнейшее свойство системы действительных чисел — ее непрерывность. Система действительных чисел — это абстрактный образ всевозможных значений непрерывно изменяющейся величины.

Таким образом, подобно арифметике целых чисел, арифметика действительных чисел имеет своим предметом реальные количественные отношения непрерывных величин, которые она изучает в их общем виде, в полном отвлечении от всякой конкретности. Именно потому, что в понятии действительного числа выделяется общее всем непрерывным величинам, оно имеет столь широкое применение: значения разных величин, будь то длина, вес, сила электрического тока, энергия и т. п., выражаются числами, а зависимость, связь между величинами изображаются, как зависимость между их численными значениями.

Для того, чтобы общее понятие о действительных числах могло служить основанием математической теории, нужно дать формально математическое их определение. Это можно сделать разными путями, но, пожалуй, естественнее всего исходить из самого процесса измерения величин, который как раз и служил практическим источником обобщения понятия о числе. Мы будем говорить о длине отрезков, но читатель легко заметит, что точно так же можно было бы рассуждать о любых других величинах, допускающих неограниченное деление.

Пусть мы хотим измерить отрезок AB посредством принятого за единицу отрезка CD (рис. 1). Откладываем отрезок CD на AB, например от точки A, пока он еще

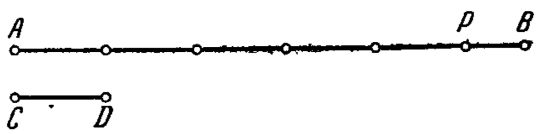


Рис. 1.

укладывается в AB. Пусть он отложился n_0 раз. Если после этого от отрезка AB остался еще остаток PB, то делим отрезок CD на десять частей и измеряем остаток этими десятыми долями. Пусть в остатке отложилось n_1 десятых долей. Если и после этого есть остаток, то делим нашу мерку еще раз на десять частей, т. е. делим CD на сотые доли, и повторяем ту же операцию, и т. д. Либо процесс измерения кончится, либо он будет продолжаться. Но во всяком случае мы будем получать, что в отрезке AB укладывается n_0 целых отрезков CD, n_1— десятых, n_2— сотых и т. д. Словом, мы будем получать отношение AB к CD все с большей точностью: до десятых, до сотых и т. д. Само отношение представится, стало быть, десятичной дробью с n_0 целыми, n_1 десятыми и т. д.

$$\frac{AB}{CD} = n_0,\, n_1 n_2 n_3 \ldots$$

Эта дробь может быть бесконечной, что означает возможность неограниченного уточнения измерения.

Итак, отношение отрезков (и величин вообще) представимо всегда десятичной дробью (конечной или бесконечной). Однако в десятичной дроби уже нет следов самой конкретной величины. Поэтому она дает как раз отвлеченное отношение, т. е. действительное число. Действительное число формально так и можно определить как конечную или бесконечную десятичную дробь[1].

Для того, чтобы довести дело до конца, нужно еще определить действия (сложение и т. п.) над десятичными дробями. Это делается таким образом, чтобы определяемые действия над десятичными дробями отвечали действиям над величинами. Так, при сложении отрезков их длины складываются, т. е. длина отрезка $AB + BC$ равна сумме длин AB и BC. В определении действий над действительными числами есть та трудность, что числа эти представляются, вообще говоря, *бесконечными* десятичными дробями, тогда как обычные известные правила действий относятся к конечным десятичным дробям. В связи с этим строгое определение действий с бесконечными дробями дается следующим путем. Пусть, например, нам нужно сложить два числа a и b. Берем соответствующие десятичные дроби с точностью до данного знака, например до миллионных, и складываем их. Тогда получим сумму $a + b$ с соответствующей точностью (до двух миллионных, так как ошибки от a и b могут сложиться). Таким образом, мы можем определить сумму двух чисел с *любой степенью точности*, и в этом смысле их сумма оказывается вполне определенной, хотя на каждом этапе вычисления она известна лишь с некоторой точностью. Это, однако, отвечает существу дела, потому что каждая величина a и b также измеряется с некоторой точностью, и точное значение, представляемое бесконечной десятичной дробью, получается как результат неограниченно продолжаемого возможного уточнения значения величины.

Отношение «больше», «меньше» может быть затем определено через сложение: $a > b$, если существует такое c, что $a = b + c$ (речь идет о положительных числах).

Непрерывность ряда действительных чисел выражается в том, что если числа a_1, a_2, ... возрастают, а b_1, b_2, ... убывают, оставаясь при этом всегда больше чисел a_i, то между теми и другими числами всегда есть еще какое-либо число c. Наглядно это

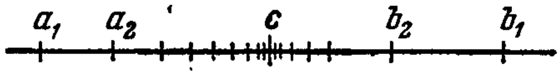


Рис. 2.

изображается на прямой, если ее точки по известному правилу сопоставить с числами (рис. 2). Здесь ясно видно, что наличие числа c и соответствующей ему точки как раз означает отсутствие разрыва в ряду чисел, т. е. непрерывность ряда чисел.

4. Уже на примере взаимодействия арифметики и геометрии можно видеть, что развитие математики происходит в процессе борьбы многих сплетающихся в ней противоположностей: конкретного и абстрактного, частного и общего, формального и содержательного, конечного и бесконечного, прерывного и непрерывного и т. д.

Попробуем, например, проследить противоположность конкретного и абстрактного в самом создании понятия действительного числа. Как мы видели, действительное число отражает *бесконечно уточняемый* процесс измерения или, в несколько ином понимании, *абсолютно*, *бесконечно* точ-

[1] Дроби с девяткой в периоде при этом не рассматриваются, они отождествляются с соответствующей дробью без девяток в периоде по известному правилу, ясному из примера: 0,139999... = 0,14000...

ное значение величины. Это соответствует тому, что в геометрии рассматривают идеально точные формы и размеры тел, вовсе отвлекаясь от подвижности и некоторой неопределенности реальных форм и размеров конкретных предметов. Выше мы рассуждали об измерении именно идеального отрезка.

Однако идеально точные геометрические формы и абсолютно точные значения величин представляют собою абстракции. Никакой конкретный предмет не имеет абсолютно точной формы, как никакая конкретная величина не только не измерима абсолютно точно, но и *не имеет* абсолютно точного значения. Длины линеек, например, не имеют смысла, если их уточнять за пределы атомных масштабов. Всегда за известными пределами количественного уточнения происходит качественное изменение величины, и она теряет вообще свой первоначальный смысл. Например, давление газа нельзя уточнять за пределы силы удара одной молекулы; электрический заряд перестает быть непрерывным, когда уточняется до заряда электрона и т. п. Ввиду отсутствия в природе предметов идеально точной формы, утверждение о том, что отношение диагонали квадрата к стороне равно $\sqrt{2}$, не только нельзя абсолютно точно вывести из непосредственного измерения, но оно и не имеет абсолютно точного смысла ни для какого конкретного реального квадрата.

Вывод о несоизмеримости диагонали и стороны квадрата вытекает, как мы видели, из теоремы Пифагора. Это — *теоретический* вывод в развитие данных опыта; он есть результат применения логики к исходным, взятым из опыта, посылкам геометрии.

Таким образом, понятие о несоизмеримых отрезках и тем более о действительном числе не есть простое, непосредственное отражение опытных фактов, но идет как бы дальше их. Это, однако, понятно. Действительное число не отражает какую-либо данную конкретную величину, а величину вообще в отвлечении от всякой конкретности, иными словами, оно отражает *общее* в реальных частных величинах. Это общее состоит, в частности, в том, что значение величин вообще можно уточнять, и если мы отвлеклись от конкретных величин, то граница этого возможного уточнения, зависящая от конкретной природы величины, становится неопределенной и исчезает.

Таким образом, *математическая* теория величин, рассматривающая величины в отвлечении от их индивидуальной природы, неизбежно должна рассматривать возможность неограниченного уточнения значения величины и тем самым *должна* приводить к понятию действительного числа. Вместе с тем, отражая только общее в разных величинах, математика не учитывает особенностей каждого отдельного случая. «Всякое отдельное, — как отмечает В. И. Ленин, — неполно входит в общее...» [1]

Выделяя общие свойства, математика рассматривает выработанные ею четко разграниченные абстракции независимо от реальных

[1] В. И. Ленин. Философские тетради. Госполитиздат, 1947, стр. 329.

границ их применимости. И это происходит именно потому, что границы эти не обладают той же общностью; они зависят от конкретных свойств рассматриваемых явлений, от качественной смены этих явлений. Поэтому, применяя математику, необходимо проверять обоснованность самого применения той или иной теории. Так, рассматривать вещество как непрерывное и описывать его свойства непрерывными величинами допустимо лишь тогда, когда можно отвлечься от его атомного строения, а это возможно лишь при известных условиях в некоторых пределах.

Несмотря на это, действительные числа представляют собой проверенное и мощное средство математического исследования реальных непрерывных величин и процессов. Теория их обоснована практикой — необозримым полем применений в физике, технике, химии. Стало быть, практика доказывает, что понятие дейсвительного числа правильно отражает общие свойства величин. Но эта правильность не безгранична; теорию действительных чисел нельзя рассматривать как нечто абсолютное, допускающее неограниченное абстрактное развитие в полном отвлечении от действительности. Само понятие о действительном числе продолжает развиваться и фактически далеко еще не является абсолютно законченным.

5. Роль еще одной из названных выше противоположностей — противоположности дискретного и непрерывного можно также проследить на примере развития понятия числа. Мы уже видели, что дроби возникли из деления непрерывных величин.

На тему о делении существует чрезвычайно поучительный шуточный вопрос. Бабушка купила три картофелины и должна разделить их поровну между двумя внуками. Как быть? Ответ гласит: нужно сделать пюре.

Эта шутка, однако, вскрывает самую суть дела. Отдельные предметы неделимы в том смысле, что разделенный предмет почти всегда перестает быть тем, что он есть, как это ясно из примеров «трети человека» или «трети стрелы». Напротив, непрерывные и однородные величины или вещи легко делятся и складываются, не теряя своей сущности. Пюре как раз представляет прекрасный пример такой однородной вещи, которая хотя и не разделена, но зато легко делится практически на сколь угодно мелкие доли. Длины, площади, объемы обладают тем же свойством. Непрерывное по самому понятию и есть то, что не разделено в действительности, но неограниченно делимо в возможности.

Мы сталкиваемся, таким образом, с двумя противоположностями: с одной стороны, неделимые, отдельные, как говорят, дискретные предметы, с другой — вполне делимые, но не разделенные на части, а непрерывные вещи. Конечно, эти противоположности всегда соединены, ибо нет ни абсолютно неделимых, ни совершенно непрерывных предметов. Однако эти стороны предметов не только реальны, но часто в одних случаях решающей оказывается одна сторона, в других — другая.

Математика, отвлекая формы от их содержания, тем самым резко разделяет и эти формы — дискретное и непрерывное.

Математическим образом отдельного предмета служит единица, а математическим образом совокупности дискретных предметов служит сумма единиц. Это, так сказать, образ чистой дискретности, очищенной от всего остального, дискретность в ее чистом виде. Основным, исходным математическим образом непрерывности служит непрерывность геометрической фигуры, в простейшем случае — прямой линии.

Перед нами, стало быть, две противоположности: дискретность и непрерывность, и их отвлеченные математические образы: целое число и геометрическая протяженность. Измерение есть соединение этих противоположностей: непрерывное измеряется отдельными единицами. Но неделимыми единицами обойтись нельзя; приходится вводить дробные доли исходной единицы. Так возникают дробные числа; понятие числа развивается именно в результате соединения указанных противоположностей.

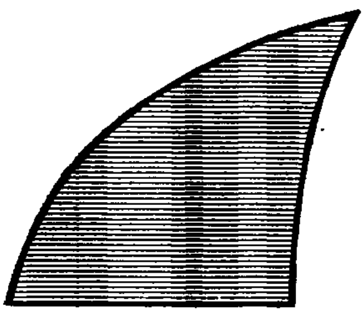


Рис. 3.

Дальше, на более абстрактной ступени появляется понятие о несоизмеримых отрезках и, как следствие, понятие действительного числа как отвлеченный образ неограниченно точного значения величины. Однако это понятие сложилось не сразу, и долгий путь его развития шел через борьбу тех же противоположностей дискретного и непрерывного.

Как уже было сказано, Демокрит представлял фигуры составленными из атомов и тем сводил непрерывное к прерывному. Однако открытие несоизмеримых отрезков заставило отказаться от такого представления. После этого непрерывные величины уже не составлялись из отдельных элементов — атомов или точек. Эти величины не выражались числами, так как иных чисел, кроме целых и дробных, в то время не знали.

Противоречие прерывного и непрерывного выступило в математике с новой силой в XVII в., когда закладывались основы дифференциального и интегрального исчисления. Здесь речь шла о бесконечно малых. В одних представлениях они мыслились как действительные, «актуально» бесконечно малые, «неделимые» частицы непрерывной величины, подобно атомам Демокрита, но теперь число их считалось бесконечно большим. Вычисление площадей и объемов — интегрирование — мыслилось как суммирование бесконечного числа этих бесконечно малых частиц. Площадь, например, понималась как «сумма линий, из которых она составляется» (рис. 3). Стало быть, непрерывное опять как бы сводилось к дискретному, но уже более сложным образом, на более высокой ступени. Но этот взгляд оказался неудовлетворительным, и, в противовес ему, в основном от Ньютона пошло представление о *непрерывных* переменных, о бесконечно малых как неограниченно убывающих *переменных* величинах.

Эта концепция взяла верх, когда в первой половине XIX в. была создана строгая теория пределов. Теперь отрезок не составлялся из точек или «неделимых», но понимался как протяженность, как непрерывная среда, где можно лишь фиксировать отдельные точки, отдельные значения переменной величины. Математики так и говорили тогда о «протяженности». В единстве прерывного и непрерывного непрерывность стала опять господствующей.

Однако развитие анализа потребовало дальнейшего уточнения теории переменных величин и прежде всего общего определения действительного числа как любого возможного значения переменной величины. Тогда в 70-х годах прошлого столетия возникла теория действительных чисел, которая представляет отрезок как множество точек и соответственно промежуток изменения переменной — как множество действительных чисел. Непрерывность опять стала составляться из отдельных дискретных точек, и свойства непрерывности стали выражаться в строении совокупности составляющих ее точек. Эта концепция привела к громадным успехам математики и стала господствующей. Все же и в ней обнаружились свои глубокие трудности, которые породили попытки вернуться на новой ступени к представлению о чистой непрерывности. Намечаются также иные пути переделки представления об отрезке как множестве точек. Возникают новые точки зрения на понятия числа, переменной, функции. Развитие теории продолжается, и нужно ждать его дальнейших шагов.

6. Взаимодействие арифметики и геометрии играло роль не только в создании понятия действительного числа. То же взаимодействие геометрии и арифметики, или, точнее, уже алгебры, сказалось также в утверждении в математике понятий отрицательных и комплексных чисел (т. е. чисел вида $a + b\sqrt{-1}$). Отрицательные числа изображаются точками на прямой слева от точки, которая сопоставляется с нулем. Комплексные же числа изображают точками на плоскости. Именно это геометрическое представление укрепило в математике мнимые числа, которые до того оставались непонятными.

Понятие величины развивалось дальше: появились, например, векторные величины, которые изображают направленными отрезками, и другие, еще более общие величины (тензоры), в которых опять алгебра соединяется с геометрией.

Соединение различных математических теорий всегда играло и играет большую, порой решающую роль. Мы увидим это дальше на примерах возникновения аналитической геометрии, дифференциального и интегрального исчисления, теории функций комплексной переменной, новейшего так называемого функционального анализа и других теорий. В самой теории чисел, т. е. в учении о целых числах, с большим успехом применяются методы, связанные с непрерывностью: анализ бесконечно малых и геометрия, что породило обширные главы этой теории, называемые «аналитической теорией чисел» и «геометрией чисел».

С известной точки зрения в основе математики можно видеть сочетание понятий, исходящих из геометрии и арифметики, — общих понятий непрерывности и алгебраической операции (как обобщения арифметического действия). Но здесь мы не можем говорить об этих трудных теориях. Целью этого параграфа было создать хотя бы самое общее представление о взаимодействии понятий, о единстве и борьбе противоположностей в математике на примере взаимодействия арифметики и геометрии, на примере развития понятия числа.

§ 5. ЭПОХА ЭЛЕМЕНТАРНОЙ МАТЕМАТИКИ

1. Развитие математики не сводится к простому накоплению новых теорем, но включает существенные, можно сказать, качественные изменения математики. Эти качественные изменения происходят, однако, не в порядке ломки и отмены существующих теорий, но в порядке их углубления и обобщения, в порядке появления новых обобщающих теорий, подготовленных предшествующим развитием.

С самой общей точки зрения в истории математики можно отметить четыре основных, качественно различных этапа. Конечно, точное разграничение этих этапов невозможно, потому что существенные черты каждого следующего из них складывались более или менее постепенно, но различие этих этапов и переходы между ними обозначаются вполне отчетливо.

Первый этап (или период) — это период зарождения математики как самостоятельной и чисто теоретической науки. Он тянулся с древнейших времен и закончился к V в. до н. э., если не раньше, когда в Греции сложилась, наконец, «чистая» математика с ее логической связью теорем и доказательств (в V в. до н. э. появились, в частности, систематические изложения геометрии, например, «Начала» Гиппократа Хиосского). Этот период был периодом формирования арифметики и геометрии, и мы уже достаточно подробно его рассмотрели. Тогда математика складывалась как *непосредственно* связанная с практикой совокупность отдельных правил, выведенных из опыта. Эти правила не образовывали еще единой логически связанной системы. Теоретический характер математики с её логическим доказательством теорем складывался очень медленно, по мере накопления материала. Арифметика и геометрия не были разделены, но тесно переплетались друг с другом.

Второй период можно характеризовать как период элементарной математики, математики постоянных величин; основные, простейшие его результаты составляют теперь содержание школьного курса. Этот период продолжался около двух тысяч лет и закончился в XVII в. с возникновением «высшей» математики. На этом периоде мы подробнее остановимся в этом параграфе. Следующие параграфы будут посвящены третьему и четвертому периодам — эпохе создания и развития анализа и периоду современной математики.

2. Период элементарной математики можно в свою очередь разделить на две части, различающиеся их основным содержанием: период развития геометрии (до II в. н. э.) и период преимущественного развития алгебры (от II до XVII в.). По историческим же условиям он делится на три периода, которые можно назвать «греческим», «восточным» и «европейским эпохи Возрождения». Греческий период совпадает по времени с общим расцветом греческой культуры; восходя началом к VII в. до н. э., он достиг своей вершины в III в. до н. э., во времена величайших геометров древности Эвклида, Архимеда, Аполлония, и закончился к VI в. н. э. Математика, особенно геометрия, достигла в Греции удивительного расцвета; нам известны имена и результаты очень многих греческих математиков, хотя до нас дошли только немногие из их подлинных сочинений. При этом стоит заметить, что Рим, достигший расцвета к I в. н. э., не дал в математике ничего, в то время как в порабощенной им Греции наука еще процветала.

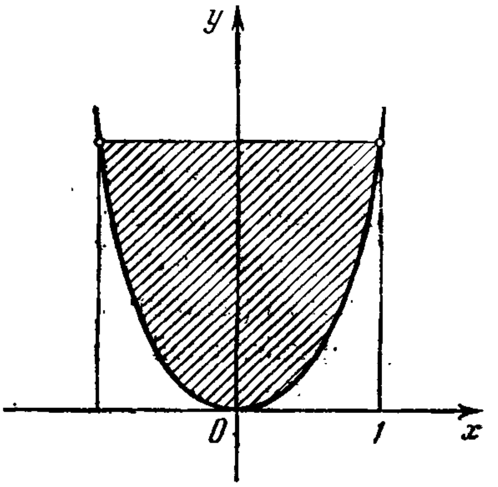


Рис. 4.

Греки не только развили и привели в стройную систему элементарную геометрию в том объеме, в каком она дана в «Началах» Эвклида и преподается теперь в школах, но достигли гораздо больших результатов. Так, они изучили конические сечения: эллипс, гиперболу и параболу; доказали некоторые теоремы, относящиеся к началам так называемой проективной геометрии; разработали, руководствуясь потребностями астрономии, геометрию на сфере (I в. н. э.), а также начала тригонометрии и вычислили первые таблицы синусов (Гиппарх — II век до н. э. и Клавдий Птолемей — II в. н. э.[1]); определили ряд площадей и объемов сложных фигур, например, Архимед определил площадь сегмента параболы, доказав, что она составляет $^2/_3$ площади прямоугольника, содержащего этот сегмент (рис. 4). Грекам была известна даже, например, такая теорема, что из всех тел с данной площадью поверхности наибольший объем имеет шар, но доказательства ее не сохранилось, и едва ли греки владели полным ее доказательством, столь оно трудно; оно было впервые найдено в XIX в. посредством интегрального исчисления.

В области арифметики и начал алгебры греки также дали немало. Они, как уже было упомянуто раньше, положили начало теории чисел. Сюда относятся, например, их исследования о простых числах (теорема Эвклида о существовании бесконечного множества простых чисел и «ре-

[1] Птолемей широко известен как автор системы, по которой центром мира считалась Земля и движение светил описывалось, как происходящее вокруг Земли. Эта система была опровергнута Коперником.

шето Эратосфена» для нахождения простых чисел), а также решение уравнений в целых числах (Диофант, около 246—330 гг. н. э.).

Мы говорили уже, что греки открыли иррациональные величины, но рассматривали их геометрически, как отрезки. Поэтому задачи, которые мы теперь рассматриваем алгебраически, греки рассматривали геометрически. Так они решали квадратные уравнения и преобразовывали иррациональные выражения. Например, уравнение, которое мы пишем теперь в виде $x^2 + ax = b^2$, читалось так: найти такой отрезок x, что если к построенному на нем квадрату приложить прямоугольник, построенный на том же отрезке и данном отрезке a, получим прямоугольник, равновеликий данному квадрату. Господство геометрии продолжалось долгое время после греков. Греки знали также способы извлечения квадратного и кубического корней, свойства арифметических и геометрических прогрессий.

Таким образом, у греков был уже большой материал из современной элементарной алгебры, но не хватало главного: отрицательных чисел и нуля, иррациональных чисел в отвлечении от всякой геометрии и, наконец, развитой системы буквенных обозначений. Впрочем, Диофант уже употреблял буквенные обозначения для неизвестной и ее степеней, а также специальные знаки сложения, вычитания, равенства, поэтому он писал алгебраические уравнения, однако еще только с конкретными численными коэффициентами.

В геометрии греки вплотную подошли к «высшей» математике: Архимед — к интегральному исчислению в вычислении площадей и объемов, Аполлоний — к аналитической геометрии в своих исследованиях о конических сечениях. Он фактически дает уравнения этих кривых[1], но выражает их геометрическим языком. Однако у них не было еще ни общих понятий произвольной постоянной и переменной величины, ни той необходимой формы — буквенных обозначений алгебры, которые появились совсем в другую эпоху и которые только и могли превратить их исследования в источник новых теорий, входящих уже в высшую математику. Создатели этих теорий тысячу лет спустя отправлялись в большой мере от наследия греческих ученых. Сочинение Декарта «Геометрия» (1637), кладущее начало аналитической геометрии, как раз начинается с разбора задач, оставленных греками.

Таков общий закон. Старые теории, порождая новые и глубокие задачи, как бы перерастают сами себя и требуют тогда для дальнейшего развития новых форм, новых идей. Эти новые формы и идеи для своего возникновения могут требовать иных условий. В античном обществе не было и не могло быть условий для перехода к высшей математике; они наступили с развитием естествознания в новое время, а это развитие в свою

[1] Он дает «уравнения» конических сечений, отнесенных к вершине. Например, «уравнение» параболы $y^2 = 2px$ он формулирует так: квадрат со стороной y равновелик прямоугольнику со сторонами $2p$ и x. (Конечно, вместо обозначений p, x, y, у него фигурировали соответствующие отрезки.)

очередь было обусловлено в XVI—XVII вв. новыми потребностями техники и промышленности и было связано, таким образом, с зарождением и развитием капитализма.

Греки как бы исчерпали возможности элементарной геометрии, и этим надо объяснить тот факт, что блестящий прогресс геометрии иссяк к началу нашей эры и сменился развитием тригонометрии и алгебры в работах Птолемея, Диофанта и др. Как раз работы Диофанта можно считать началом того периода, когда ведущей становится алгебра. Но шедшее к своему закату античное общество уже не могло двигать науку дальше в этом новом направлении.

Надо отметить, что уже за несколько веков до этого в Китае арифметика достигла высокого уровня. Китайскими учеными во II—I вв. до н. э. описываются правила арифметического решения системы трех уравнений первой степени. При этом впервые в истории используются отрицательные коэффициенты и формулируются правила действия с отрицательными количествами. (Сами решения они, однако, искали только положительные, так же, как это позже делал Диофант.) В тех же книгах уже фигурирует способ извлечения квадратных и кубических корней.

3. С концом греческой науки в Европе наступил научный застой, и центр развития математики переместился в Индию, Среднюю Азию и арабские страны[1]. Здесь на протяжении тысячи лет, с V по XV в., математика развилась главным образом в связи с потребностями вычислений, в частности для астрономии; математики Востока по большей части были также астрономами. Ими, правда, почти ничего не было прибавлено значительного к греческой геометрии; в этой науке они лишь сохранили для последующих времен творения греков. Зато индийскими, арабскими и среднеазиатскими математиками были достигнуты громадные успехи в области арифметики и алгебры[2].

Индийцы, как уже говорилось в § 2, изобрели современную систему счисления. Они ввели также отрицательные числа, связывая противоположение положительных и отрицательных чисел с противоположением имущества и долга или двух направлений на прямой. Они, наконец, начали оперировать с иррациональными количествами так же, как с рациональными, без геометрического их представления, в отличие от греков.

[1] Для ориентировки во времени укажем периоды жизни некоторых выдающихся математиков Востока. Индийцы: Ариабхата — род. ок. 476 г.; Брамагупта — ок. 598—660 гг.; Бхаскара — XII в.; хорезмийцы: Аль-Хорезми — IX в.; Аль-Бируни — 973—1048 гг.; работавший в Азербайджане Насирэддин Туси — 1201—1274 гг.; работавший в Самарканде Гиясэддин Джемшид — XV в.

[2] Нужно иметь в виду, что связывать развитие математики того времени главным образом с арабами неверно. Термин «арабская» математика обусловлен преимущественно тем, что очень многие ученые Востока писали на арабском языке, распространявшемся вместе с арабскими завоеваниями.

У них были также специальные обозначения для алгебраических действий, включая извлечение корня. Именно благодаря тому, что индийские и среднеазиатские ученые не смутились различием иррациональных и рациональных количеств, они смогли преодолеть «засилие» геометрии, характерное для греческой математики, и открыли путь развитию настоящей алгебры, свободной от тяжеловесной геометрической оболочки, в которую она была втиснута греками.

Великий поэт и математик Омар Хайям (ок. 1048—1122) и Насирэддин Туси явно указывали, что каждое отношение величин, все равно соизмеримых или несоизмеримых, может быть названо числом; тем самым у них мы находим уже то общее определение числа (как рационального, ак и иррационального), которое в формулировке Ньютона было приведено выше, в § 4. Величие этих достижений становится особенно ясным, если заметить, что полное признание отрицательных и иррациональных чисел всеми европейскими математиками было достигнуто очень нескоро, даже после того, как в Европе началось возрождение математики. Еще, например, знаменитый французский математик Виет (1540—1603), которому алгебра многим обязана, избегал отрицательных чисел, а в Англии протесты против отрицательных чисел раздавались даже в XVIII в. Числа эти считались нелепыми, так как они меньше нуля, т. е. «меньше чем ничто». Теперь они стали привычными хотя бы в виде отрицательной температуры; все у нас читают в газетах и понимают, что означает: «температура в Москве —8°».

Самое слово «алгебра» происходит от названия сочинения хорезмского математика и астронома Мухаммед ибн Муса аль-Хорезми (Мухаммеда сына Мусы из Хорезма), жившего в IX в.; его сочинение по алгебре называлось «Альджебр альмукабала», что означает «восстановление и противоположение». Под восстановлением — альджебр — понималось перенесение отрицательных членов в другую часть уравнения, а под противоположением — альмукабала — отбрасывание в обеих частях равенства равных членов.

Арабское слово «альджебр» при переводе на латинский язык превратилось в algebra, а альмукабала была отброшена: так появилось самое название «алгебра» [1].

Кстати, происхождение этого названия вполне отвечает содержанию самой науки. Алгебра в своей основе есть учение об арифметических действиях, рассматриваемых формально в общем виде, отвлекаясь от конкретных чисел. Ее задачи составляют в первую очередь формальные правила преобразования выражений и решения уравнений. Аль-Хорезми поставил заглавием своей книги как раз названия некоторых общих формальных правил и тем выразил дух алгебры.

[1] Стоит заметить еще, что математический термин «алгорифм», означающий метод, правила вычисления, происходит от имени того же Аль-Хорезми.

Позже Омар Хайям дал определение алгебры как науки о решении уравнений. Это определение сохраняло свое значение до конца прошлого века, когда наряду с теорией уравнений в алгебре сложились новые направления, существенно изменившие ее лицо, но не изменившие ее общего духа как общего учения о формальных действиях.

Среднеазиатские математики нашли методы извлечения корней и приближенного решения ряда уравнений, общую формулу «бинома Ньютона», хотя выражали ее словами, и др. Они сильно продвинули тригонометрию, приведя ее в систему, и вычислили очень точные таблицы синусов. Эти таблицы вычислил в связи с потребностями астрономии работавший у знаменитого узбекского астронома Улугбека математик Гиясэддин (около 1427), который изобрел также десятичные дроби за 150 лет до их вторичного изобретения в Европе.

Словом, в течение средних веков в Индии и в Средней Азии сложились почти полностью современная десятичная система счисления (включая дроби), элементарная алгебра и тригонометрия. В тот же период начали проникать в соседние страны достижения китайской науки, где около VI в. были известны приемы решения простейших неопределенных уравнений, приближенные вычисления в геометрии и первые приемы приближенного решения уравнений третьей степени. Из материала школьного курса алгебры к XVI в. не хватало, пожалуй, только логарифмов и мнимых чисел. Кроме того, не существовало еще системы буквенных обозначений: содержание алгебры обгоняло ее форму. Однако эта форма была необходима: отвлечение от конкретных чисел и формулирование общих правил требовали соответствующего способа выражения, нужно было обозначать *любые* числа и действия над ними. Алгебраическая символика является необходимой формой, отвечающей содержанию алгебры. Как в глубокой древности, чтобы оперировать с целыми числами, нужно было выработать для них обозначения, так и теперь, чтобы оперировать с произвольными числами и давать для них общие правила, нужно было выработать соответствующие обозначения. Эта задача решалась со времен греков, и ее решение было завершено лишь в XVII в., когда в трудах Декарта и других сложилась, наконец, современная система обозначений.

4. При возрождении наук европейцы учились у арабов и знакомились с греческой наукой по арабским переводам. Книги Эвклида, Птолемея, Аль-Хорезми в XII в. впервые перевели с арабского на латинский — общий научный язык Западной Европы того времени. В то же время в борьбе с прежней системой счета, идущей от греков и Рима, постепенно укрепляется в Европе индийское счисление, заимствуемое у арабов.

Только в XVI в. европейская наука, наконец, впервые превзошла достижения своих предшественников. Так, итальянцы Тарталья и Феррари решили в общем виде: первый — кубическое, второй — уравнение четвертой степени (см. главу IV). (Заметим, что, хотя эти выводы не проходят в школе, они по уровню применяемых методов принадлежат к элементарной алгебре. К высшей алгебре нужно относить общую теорию уравнений.)

В этот же период впервые начинают оперировать с мнимыми числами (пока чисто формально, без какого-либо реального обоснования, которое выясняется гораздо позже, в начале XIX в.). Вырабатываются также современные алгебраические обозначения и, в частности, появляются (у Виета в 1591 г.) буквенные обозначения не только неизвестных, но и данных чисел: «a», «b» и т. п. Во всей этой работе по развитию алгебры принимает участие много математиков. Тогда же, кстати, появляются в Европе десятичные дроби (их изобретает нидерландский ученый Стевин и пишет о них в 1585 г.).

Наконец, Непер в Англии изобретает логарифмы как пособие для астрономических вычислений и сообщает о них в 1614 г., а Бригг вычисляет первые десятичные таблицы логарифмов, которые появляются в 1624 г.[1]

Тогда же появляются в Европе «теория соединений» и общая формула «бинома Ньютона»[2]; прогрессии были уже известны раньше. Таким образом, построение элементарной алгебры завершается. Вместе с этим в начале XVII в. заканчивается весь период математики постоянных величин, элементарной математики, которую теперь с небольшими добавлениями изучают в школе; арифметика, элементарная геометрия, тригонометрия, элементарная алгебра сложились к тому времени во всем существенном. Дальше следовал переход к высшей математике — математике переменных величин.

Не следует, однако, думать, будто на этом кончилось развитие элементарной математики. Оно продолжается и, например, в элементарной геометрии постоянно появлялись и появляются новые результаты. Более того, именно благодаря дальнейшему развитию математики мы более ясно осознали сущность самой элементарной математики. Однако руководящую роль в математике теперь уже приобрели понятия переменной величины, функции, предела. Задачи, идущие из элементарной математики, теперь часто не только освещаются и решаются с помощью этих понятий высшей математики и связанных с ними методов, но они подчас и не разрешимы элементарными методами. В той же связи с понятиями и методами высшей математики задачи, идущие от элементарной математики,

---

[1] Интересно отметить, что Непер определял логарифм не так, как теперь, когда говорят, что в формуле $x = a^y$ число y есть логарифм x при основании a. Такое определение логарифмов появилось позже. Определение Непера было связано с понятиями переменной величины и бесконечно малых и сводилось к тому, что логарифм x есть такая функция $y = f(x)$, скорость роста которой обратно пропорциональна x, т. е. $y' = \frac{c}{x}$ (см. главу II). Таким образом, по существу исходным было дифференциальное уравнение, определяющее логарифм, хотя не были изобретены еще дифференциалы.

[2] Эта формула носит имя Ньютона не потому, что он впервые открыл ее, а

служат и теперь источником более общих результатов и даже теорий. Примеры тому представляет уже упомянутая теория правильных систем фигур, или задачи теории чисел, элементарные по формулировке, но вовсе не элементарные по методам их решения, о чем читатель может подробнее узнать в главе X (том 2).

§ 6. МАТЕМАТИКА ПЕРЕМЕННЫХ ВЕЛИЧИН

1. К XVI в. исследование движения стало центральной задачей естествознания. К исследованию движения, исследованию различных процессов изменения и зависимостей между изменяющимися величинами естественные науки были подведены запросами практики и всем развитием самих этих наук.

Как отражение общих свойств изменяющихся величин и зависимостей между ними в математике возникли понятия переменной величины и функции, и это кардинальное расширение предмета математики определило переход к новому ее этапу — математике переменных величин.

Закон движения тела по данной траектории, например по прямой, определяется тем, как нарастает со временем пройденный телом путь.

Так, Галилей (1564—1642) открыл закон падения, установив, что путь, проходимый падающим телом, нарастает пропорционально квадрату времени. Это выражается известной формулой

$$s = \frac{gt^2}{2} \tag{1}$$

(где $g \approx 9{,}81$ м/сек^2).

Вообще закон движения задает путь, пройденный за время t. Здесь время t и путь s — две переменные: «независимая» и «зависимая», а тот факт, что каждому t отвечает определенный путь s, означает, что путь s есть функция времени t.

Математические понятия переменной и функции представляют собой не что иное, как абстрактное обобщение конкретных переменных величин (как время, путь, скорость, угол поворота, зачерчиваемая площадь и т. д.) и конкретных зависимостей между ними (как зависимость пути от времени и т. д.). Как понятие действительного числа есть отвлеченный образ значения любой величины, так «переменная» есть отвлеченный образ изменяющейся величины — величины, необходимо принимающей в рассматриваемом процессе разные значения. Математическая переменная величина x есть не что иное, как «нечто», или, лучше сказать, что угодно, что может принимать разные численные значения. Это есть, стало быть, переменная вообще; под ней можно разуметь и время, и путь, и любую другую величину.

Совершенно так же функция есть отвлеченный образ зависимости одной величины от другой. Утверждение, что y есть функция x, обозначает

потому, что он обобщил ее с целых положительных степеней n на любые дробные и иррациональные.

в математике только то, что каждому значению, которое может принять x, отвечает определенное значение y. (Функцией называют также самое соответствие или закон соответствия значений y значениям x.) Например, по закону падения, пройденный путь связан со временем падения формулой (1). Путь есть функция времени.

Энергия движущегося тела выражается через его массу и скорость по формуле

$$E = \frac{mv^2}{2}. \tag{2}$$

Для данного тела энергия E есть функция скорости v.

По известному закону количество тепла, выделяемое в проводе в единицу времени при прохождении тока, выражается формулой

$$Q = \frac{RI^2}{2}, \tag{3}$$

где I — сила тока, а R — сопротивление провода. При данном сопротивлении каждой силе тока I отвечает определенное количество тепла Q, выделяемое в единицу времени. Стало быть, Q есть функция I.

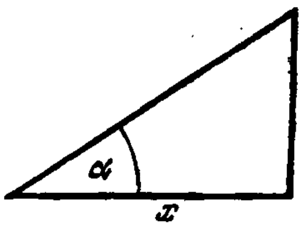


Рис. 5.

Площадь S прямоугольного треугольника с данным острым углом α и прилежащим катетом x (рис. 5) выражается формулой

$$S = \frac{1}{2} x^2 \operatorname{tg} \alpha. \tag{4}$$

При данном угле α площадь есть функция катета x.

Все формулы (1)—(4) могут быть объединены в одной

$$y = \frac{1}{2} ax^2. \tag{5}$$

Это и есть переход от конкретных переменных величин t, s, E, Q, v и т. д. к переменным вообще x и y, от конкретных зависимостей (1), (2), (3), (4) — к их общему виду (5). Если механика и теория электричества имеют дело с конкретными формулами (1), (2), (3), связывающими конкретные величины, то математическое учение о функциях имеет дело с общей формулой (5), не связывая ее ни с какими конкретными величинами.

Следующая ступень отвлечения от конкретного состоит в том, что рассматривают не данную зависимость y от x, как $y = \frac{1}{2}ax^2$, $y = \sin x$, $y = \lg x$ и т. п., а функциональную зависимость y от x вообще, выражаемую отвлеченной формулой

$$y = f(x).$$

Эта формула означает, что величина y есть вообще *некоторая* функция x, т. е. каждому значению, которое может принять x, *каким-либо* способом отвечает определенное значение y. Предметом математики становятся не только те или иные данные функции ($y = \frac{1}{2}ax^2$, $y = \sin x$ и т. п.), но *любые* (точнее: более или менее любые) функции. Эти ступени отвлечения сначала от конкретных величин, а потом и от конкретных функций аналогичны ступеням абстракции, пройденным при образовании понятия о целом числе: сначала отвлечение от конкретных совокупностей предметов приводит к понятию об отдельных числах (1, 3, 12 и т. п.), а дальнейшее отвлечение приводит к понятию любого целого числа вообще. Это обобщение есть результат глубокого взаимодействия анализа и синтеза: анализа отдельных зависимостей и синтеза выявленных их общих черт в форме новых понятий.

Область математики, посвященная изучению функций, называется анализом, математическим анализом или, чаще, анализом бесконечно малых. Последнее название связано с тем, что важным средством изучения функций является понятие о бесконечно малой величине (содержание этого понятия и его значение разъясняются в главе II).

Так как функция есть отвлеченный образ зависимости одной величины от другой, то можно сказать, что анализ имеет своим предметом зависимости между переменными величинами, но не между теми или иными конкретными величинами, а между переменными вообще, в отвлечении от их содержания. Такое отвлечение обеспечивает широту применений анализа, так как в одной формуле, в одной теореме он охватывает бесконечное число возможных конкретных случаев. Пример тому дают уже наши простые формулы (1)—(5). Здесь видна полная аналогия анализа с арифметикой и алгеброй. Все они возникли из определенных практических задач и отражают в общем, отвлеченном виде реальные количественные отношения действительности.

2. Итак, новый период в математике, начавшийся в XVII в.,— период математики переменных величин, можно определить как период появления и развития анализа. (Это третий из перечисленных выше крупных этапов развития математики.) Понятно, однако, что никакая теория не возникает в результате одного образования новых понятий, так и анализ не мог появиться из одних понятий переменной и функции. Для создания теории, и тем более целой области науки, какой является математический анализ, нужно, чтобы новые понятия, так сказать, пришли в действие, чтобы через них открывались новые взаимосвязи, чтобы они позволяли решать новые задачи.

Более того, сами новые понятия зарождаются, развиваются, уточняются, обобщаются только на основе тех задач, которые они позволяют решить, только на основе тех теорем, в которые они входят. Понятия переменной и функции не возникли сразу в готовом виде у Галилея, Декарта, Ньютона или кого-либо еще. Они зарождались у многих математиков (как, например, у Непера в связи с логарифмами), потом приняли более или менее отчетливую, но далеко не окончательную форму у Ньютона и Лейбница и уточнялись и обобщались дальше с развитием анализа. Современное определение их сложилось только в XIX в., но и оно не является *абсолютно* строгим и *совершенно* окончательным. Развитие самого понятия функции продолжается и в настоящее время.

Математический анализ создавался на материале зарождавшейся механики, на задачах геометрии и методах и задачах, идущих из алгебры.

Первым решительным шагом в создании математики переменных величин был выход в 1637 г. книги Декарта «Геометрия», где были заложены основы так называемой аналитической геометрии. Основные идеи Декарта следующие.

Пусть мы имеем, например, уравнение

$$x^2 + y^2 = a^2. \tag{6}$$

В алгебре x и y понимали как неизвестные, и так как данное уравнение не позволяет их определить, то оно не представляло для алгебры существенного интереса. Декарт же рассматривает x и y не как неизвестные, которые нужно находить из уравнения, но как *переменные*; само же уравнение выражает тогда зависимость между этими переменными. Такое уравнение в общем виде можно, перенося все члены в левую часть, записать так:

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

Далее, Декарт вводит на плоскости координаты x, y, называемые теперь декартовыми (рис. 6). Тем самым с каждой парой значений x, y со-

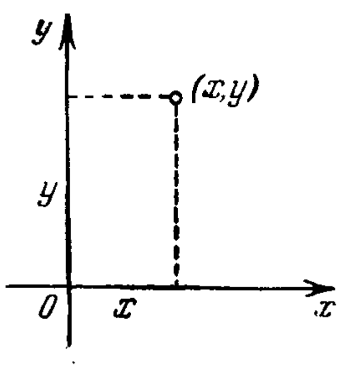


Рис. 6.

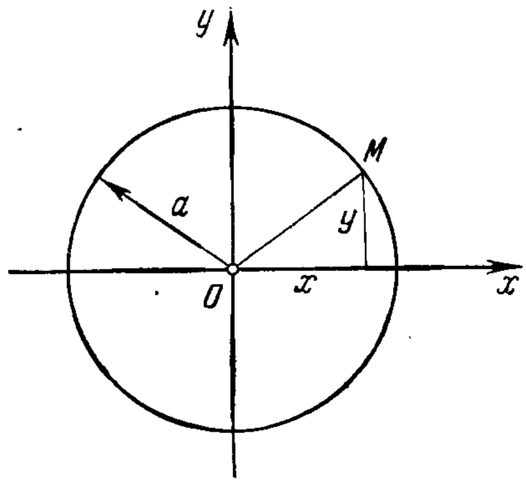


Рис. 7.

поставляется точка, и обратно: каждой точке отвечают ее координаты x, y. Благодаря этому уравнение $F(x, y) = 0$ определяет геометрическое место тех точек на плоскости, координаты которых удовлетворяют этому уравнению. Это будет, вообще говоря, некоторая линия. Например, уравнение (6) определяет окружность радиуса a с центром в начале координат. В самом деле, как видно из рис. 7, по теореме Пифагора $x^2 + y^2$ есть квадрат расстояния от начала O до точки M с координатами x и y. Поэтому уравнение (6) определяет геометрическое место тех точек, расстояние которых от начала равно a, т. е. окружность.

Обратно: геометрическое место точек, заданное тем или иным геометрическим условием, можно задавать уравнением, выражающим то же условие на языке алгебры с помощью координат. Так, например, геометрическое условие, определяющее окружность,— то, что она есть геометрическое место точек, равноудаленных от данной,— это условие на алгебраическом языке выражается уравнением (6).

Таким образом, общая задача и метод аналитической геометрии состоят в следующем: представить то или иное уравнение с двумя переменными линией на плоскости и по *алгебраическим* свойствам уравнения исследовать *геометрические* свойства соответствующей линии, и обратно: по геометрическим условиям, задающим линию, найти ее уравнение и потом опять по *алгебраическим* свойствам уравнения исследовать *геометрические* свойства этой линии. Таким путем геометрические задачи можно сводить на алгебраические и, в конце концов, на вычисление.

Содержание метода аналитической геометрии будет подробно объяснено в главе III. Сейчас мы хотим обратить внимание на то, что, как видно из нашего краткого изложения, источником этого метода служит сочетание геометрии, алгебры и общей идеи переменной величины. Главное геометрическое содержание начал аналитической геометрии составляет теория конических сечений (эллипс, гипербола, парабола). Эта теория, как уже упоминалось, была развита еще древними; выводы Аполлони уже содержали в геометрической форме уравнения конических сечений. Соединение этого геометрического содержания с алгебраической формой, подготовленной развитием математики после греков, и с общей идеей переменной величины, возникшей в связи с изучением движения, и дало аналитическую геометрию.

Если для греков конические сечения были предметом чисто математического интереса, то ко времени Декарта их изучение приобрело реальное значение для астрономии, механики и техники. Кеплер (1571—1630) открыл, что планеты движутся вокруг Солнца по эллипсам, а Галилей установил, что брошенное тело, будь то камень или пушечное ядро, летит по параболе (в первом приближении, если можно пренебречь сопротивлением воздуха). В результате вычисление разных данных, относящихся к коническим сечениям, стало насущной необходимостью. Метод Декарта решал именно эту насущную задачу. Словом, он был подготовлен предшествующим развитием математики и вызван к жизни назревшими потребностями науки и техники.

3. Следующим решающим шагом в математике переменных величин было создание Ньютоном и Лейбницем во второй половине XVII в. дифференциального и интегрального исчисления. Это и было фактическим возникновением анализа, так как предмет этих исчислений составляют свой-

ства функций самих по себе, в отличие от аналитической геометрии, предмет которой все-таки составляют еще геометрические фигуры. Ньютон и Лейбниц в действительности только завершили огромную подготовительную работу, в которой участвовали многие математики и начала которой восходят еще к приемам определения площадей и объемов, выработанным древними греками.

Мы не будем объяснять здесь содержание основных понятий дифференциального и интегрального исчисления, а также последовавших за ними теорий анализа,— это будет сделано в главах, посвященных этим теориям. Мы хотим только обратить внимание на источники дифференциального и интегрального исчисления, которыми служили главным образом новые задачи механики и достаточно старые задачи геометрии: задачи проведения касательных к кривым линиям и определения площадей и объемов. Этими задачами занимались еще древние (достаточно упомянуть Архимеда); в начале XVII в. ими занимались многие математики. Кеплер, Кавальери и другие. Однако решающим было открытие замечательной связи обоих видов задач и формулировка общего метода их решения,— это составляло заслугу Ньютона и Лейбница.

В основе открытия названной связи задач механики и геометрии лежит вытекающая из метода координат возможность графического изображения зависимости одной величины от другой, т. е. графического изображения функций. Опираясь на графическое изображение функций линиями, легко формулировать, в чем состоит связь задач механики и геометрии, лежащая в истоках дифференциального и интегрального исчисления, и в чем содержание этих исчислений.

Дифференциальное исчисление в своей основе есть метод нахождения скорости движения в любой данный момент времени, когда известна зависимость пути от времени. Эта задача решается «дифференцированием». Оказывается, она совершенно равносильна задаче проведения касательных к той линии, которая дает графическое изображение зависимости пути от времени. Скорость в момент времени t просто равна тангенсу угла наклона касательной в точке, отвечающей этому t на графике (рис. 8).

Интегральное исчисление в своей основе есть метод нахождения пройденного пути, когда известна зависимость скорости от времени (или вообще — нахождения суммарного результата действия переменной величины). Это задача, очевидно, обратная задаче дифференциального исчисления, т. е. задаче нахождения скорости: она решается «интегрированием». Оказывается, она совершенно равносильна задаче нахождения площади под кривой, которая дает графическое изображение зависимости скорости от времени. Путь, пройденный за промежуток времени от момента t_1 до t_2, просто равен площади под изображающей график скорости кривой, между прямыми, отвечающими на графике значениям t_1 и t_2 (рис. 9).

Стоит отвлечься от механической формулировки задач дифференциального и интегрального исчисления и говорить о функциях вообще, а не о зависимости пути или скорости от времени, как мы получим общее понятие о задачах дифференциального и интегрального исчисления в их чистом виде.

В основе дифференциального и интегрального исчисления, как и всего анализа в его дальнейшем развитии, помимо понятий переменной и функции, лежит также сложившееся позже понятие предела. В период

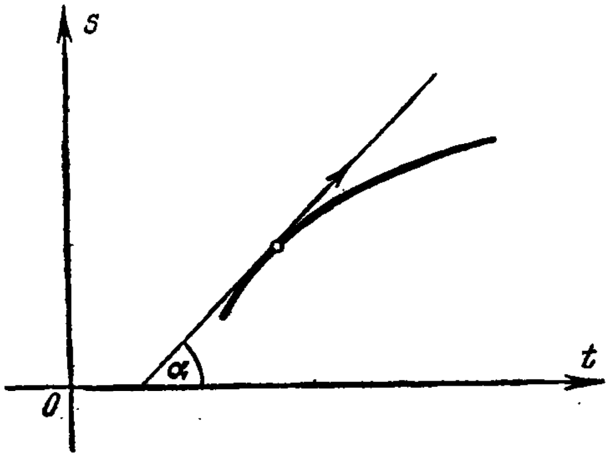


Рис. 8.

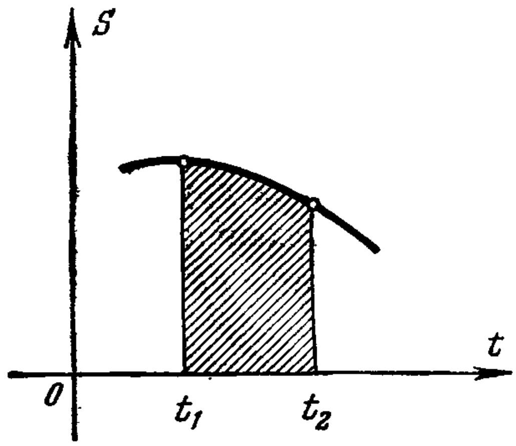


Рис. 9.

формирования анализа оно заменялось употреблением несколько расплывчатого в то время понятия бесконечно малой величины. Способы фактического вычисления скорости по закону изменения пути — «дифференцирование» и пути по скорости — «интегрирование» основаны на применении алгебры в соединении с понятием предела. Анализ возник в результате соединения этих понятий и приемов с упомянутыми задачами механики, геометрии и некоторыми другими (например, задачами на максимум и минимум). Он был насущно необходим для развития механики, в самой формулировке законов которой уже фигурируют, хотя бы в скрытом виде, понятия анализа. Второй закон Ньютона в формулировке самого Ньютона говорит, что «изменение количества движения пропорционально действующей силе». Точнее: скорость изменения количества движения пропорциональна силе. Стало быть, для того, чтобы пользоваться этим законом, нужно уметь определять скорость изменения некоторой величины, т. е. дифференцировать. Если мы будем формулировать тот же закон, говоря, что ускорение пропорционально силе, то проблема сохранится, потому что ускорение есть не что иное, как скорость изменения скорости. Само собой также понятно, что для определения закона движения, вызываемого данной переменной силой, т. е. происходящего с данным, вообще говоря, переменным ускорением, нужно уметь решать обратную задачу — находить самую величину по скорости ее изменения, т. е. нужно интегрировать. Можно сказать, что Ньютон был просто *вынужден* изобрести дифференцирование и интегрирование, чтобы иметь возможность развивать механику.

4. Вместе с дифференциальным и интегральным исчислением зародились другие отделы анализа: теория рядов (см. главу II, § 14), теория дифференциальных уравнений (главы V и VI, том 2), применение анализа к геометрии, выделившееся позже в особую область геометрии — общую теорию кривых линий и поверхностей, называемую дифференциальной геометрией (глава VII, том 2). Все эти теории также были вызваны к жизни и побуждались к развитию задачами механики, физики и техники.

Теория дифференциальных уравнений — важнейшая ветвь анализа — имеет дело с такими уравнениями, где неизвестной является уже не величина, а функция, т. е. закон зависимости одной величины от другой или от нескольких других величин. Легко понять, откуда возникают такие задачи. В механике требуется определить закон движения тела в данных условиях, а не какое-нибудь одно значение скорости или пути. В механике жидкостей требуется найти распределение скоростей по всей массе текущей жидкости, т. е. найти зависимость скорости от всех трех координат в пространстве и еще от времени. Аналогично в теории электричества и магнетизма требуется найти напряжение поля во всем пространстве, т. е. зависимость этого напряжения от тех же трех координат. И тому подобное.

Такого рода задачи постоянно возникали в механике, включая гидродинамику и теорию упругости, в акустике, в теории электричества и магнетизма, в теории тепла. Вообще с момента своего возникновения анализ развивался в самой тесной связи с развитием механики и вообще физики. Крупнейшие достижения анализа всегда были связаны с решением задач, поставленных этими науками. Начиная с Ньютона, величайшие аналитики Д. Бернулли (1700—1782) и Л. Эйлер (1707—1783), Ж. Лагранж (1736—1813) и А. Пуанкаре (1854—1912), М. В. Остроградский (1801—1861) и А. М. Ляпунов (1857—1918), как и многие другие, в своих работах, прокладывавших новые пути в анализе, исходили, как правило, из насущных задач современного им точного естествознания.

Так возникли новые теории: Эйлер и Лагранж создают в прямой связи с механикой новую ветвь анализа — так называемое вариационное исчисление (см. главу VIII, том 2), а в конце XIX в. Пуанкаре и Ляпунов, опять-таки исходя из задач механики, создают так называемую качественную теорию дифференциальных уравнений (см. главу V, том 2, § 7).

В XIX в. анализ обогатился новой важной ветвью — теорией функций комплексного переменного (см. главу IX, том 2). Зачатки этой теории имелись еще в трудах Эйлера и некоторых других математиков, но оформление ее в стройную теорию произошло к середине XIX в. и исходило в большой степени от французского математика Коши (1789—1857). Эта теория скоро достигла значительного развития и приобрела большое значение вследствие богатства своего содержания, а также потому, что она позволила глубже проникнуть в ряд законов анализа и нашла существенные приложения к решению важных задач самой математики, физики и техники.

Анализ бурно развивался, став не только центром и главной частью математики, но и проникнув в более старые ее области: алгебру, геометрию и даже теорию чисел. Алгебру стали понимать в основном как учение о функциях, выражаемых в виде многочленов от одной или нескольких переменных [1]. В геометрии стала господствовать аналитическая и дифференциальная геометрия. Наконец, еще Эйлер ввел методы анализа в теорию чисел, положив тем самым начало так называемой аналитической теории чисел, с развитием которой связаны глубочайшие достижения науки о целом числе.

Через анализ с его понятиями переменной, функции и предела во всю математику проникает идея движения, изменения и, стало быть, диалектика. Точно так же, в основном через анализ, математика испытывает на себе влияние точного естествознания и техники и сама включается в их развитие в качестве метода точной формулировки его законов и решения его задач. Как у греков математика была в основном геометрией, так, можно сказать, после Ньютона она стала в основном анализом. Конечно, анализ не поглотил всей математики целиком; в геометрии, теории чисел и алгебре всегда сохранялись специфические для них задачи и методы. Так, еще в XVII в. одновременно с аналитической геометрией зародилась другая глава геометрии — проективная геометрия, в которой господствовали чисто геометрические методы. Она имела своим источником задачи изображения предметов на плоскости (проектирование) и соответственно применяется, в частности, в начертательной геометрии.

Тогда же зародилась новая важная область математики — теория вероятностей, имеющая своим предметом закономерности, обнаруживающиеся в больших массах явлений, как серии выстрелов или бросаний монеты. Она приобрела в последнее время особое значение для физики и техники; ее расцвет, в достижении которого сыграли большую роль труды русских и советских математиков, также обусловлен идущими от естествознания и практики проблемами и применением методов анализа. Своеобразие этой теории состоит в том, что она имеет дело с законами «случайных событий», давая математические методы исследования необходимости, которая проявляется в случайности. Основы теории вероятностей будут освещены в главе XI (том 2).

5. Анализ со всеми его ответвлениями дал естествознанию и технике мощные методы решения разнообразнейших задач. Мы уже упоминали первые из них: нахождение скорости изменения какой-либо величины, когда известна зависимость самой величины от времени; определение

[1] То есть функции, например, такого вида $y=a_0x^n+a_1x^{n-1}+\ldots+a_n$. Основная задача алгебры того периода — решение уравнения $a_0x^n+a_1x^{n-1}+\ldots+a_n=0$ — означает не что иное, как поиски тех значений x, при которых функция $y=a_0x^n+a_1x^{n-1}+\ldots+a_n$ обращается в нуль. Самое существование решения — корня уравнения, т. е. основная теорема алгебры, доказывается средствами анализа (см. главу IV, § 3).

площадей криволинейных фигур и объемов тел; определение суммарного результата какого-либо процесса или суммарного действия переменной величины. Так, интегральное исчисление позволяет определить работу газа при расширении, когда давление изменяется по известному закону; то же исчисление позволяет вычислять, например, напряжение электрического поля сколь угодно сложной системы зарядов, исходя из закона Кулона, определяющего напряжение поля от одного точечного заряда, и т. п.

Далее, анализ дал метод нахождения наибольших и наименьших значений величин при тех или иных условиях. Так, с помощью анализа легко

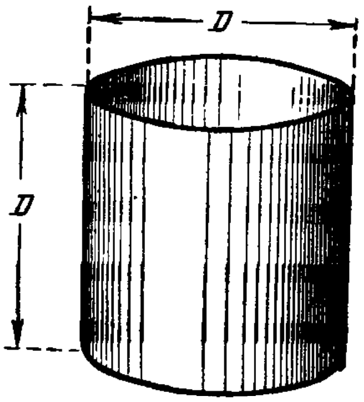


Рис. 10.

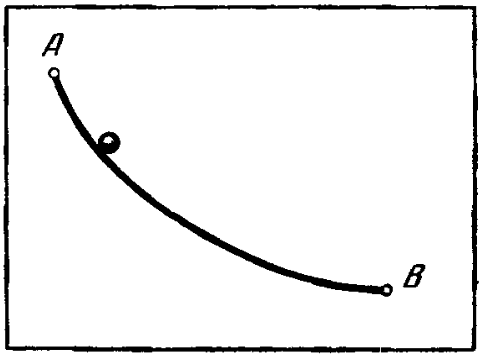


Рис. 11.

определить форму цилиндрической цистерны, которая при данном объеме имеет наименьшую поверхность и тем самым требует наименьшей затраты материала. Оказывается, это будет, когда высота цистерны равна диаметру ее основания (рис. 10). Анализ позволяет найти форму линии, по которой должно скатываться тело, чтобы в наименьшее время попасть из одной данной точки в другую (эта линия — так называемая циклоида; рис. 11).

Как решаются эти и другие подобные задачи, читатель узнает из глав II и VIII (том 2).

Анализ, точнее теория дифференциальных уравнений, дает возможность находить не просто отдельные значения переменных величин, но и неизвестные функции, т. е. законы зависимости одних величин от других. Так, мы имеем возможность, исходя из общих законов электрического тока, рассчитывать зависимость силы тока от времени при включении напряжения в любой цепи с сопротивлением, емкостью и самоиндукцией. Мы имеем возможность определять закон течения жидкости, закон распределения скоростей во всей массе жидкости при данных условиях ее течения. Мы имеем возможность вывести общие законы колебаний струн, мембран, законы распространения колебаний в различных средах: это относится к звуковым волнам, электромагнитным волнам, упругим колебаниям, распространяющимся в земле при землетрясениях и взрывах; кстати, это дает новые методы разведки полезных ископаемых и глубинного исследования грунтов. Отдельные задачи такого рода читатель встретит в главах V и VI (том 2).

Наконец, анализ дает не только способы решения тех или иных задач, он дает общие методы для самой математической формулировки количественных законов точного естествознания. Как уже сказано раньше, общие законы механики нельзя формулировать математически, не прибегая к понятиям анализа, а без такой формулировки мы не имели бы возможности решать задачи механики. Точно так же общие законы теплопроводности, диффузии, распространения колебаний, течения химических реакций, основные законы электромагнетизма и многие, многие другие просто не могут быть математически точно формулированы без понятий анализа. Только благодаря такой формулировке эти законы дают основу для применения их в разнообразнейших конкретных случаях, дают основу для точных математических выводов в отдельных задачах, касающихся теплопроводности, колебаний, растворения, электромагнитного поля, в задачах механики, астрономии, всех многочисленных разделов физики, химии, теплотехники, энергетики, машиностроения, электротехники и т. д. и т. п.

6. Подобно тому, как в истории геометрии у греков строгое и систематическое изложение, данное Эвклидом, завершало долгий путь предшествующего развития, так по мере развития анализа нарастала необходимость его обоснования, более строгого и систематического, чем то, какое давали первые творцы его действенных методов: Ньютон, Эйлер, Лагранж и другие. Создаваемый ими анализ по мере своего роста, во-первых, шел к все более и более глубоким и трудным задачам, а во-вторых, самый его объем требовал уже большей систематичности и продуманности его основ. Так, количественный рост теории необходимо порождает задачу ее лучшего обоснования, систематизации, критического разбора ее основ. «Обоснование» теории появляется как итог ее известного развития, а не исходный пункт, потому что без теории попросту неизвестно еще, что же нужно обосновывать. «Принципы,— как сказал Ф. Энгельс, — не исходный пункт исследования, а его заключительный результат»[1]. Кстати, об этом забывают некоторые современные формалисты, полагающие наиболее целесообразным излагать и даже развивать теории, исходя из аксиом, не предваренных никаким разбором того реального содержания, которое они должны суммировать. Но аксиомы сами по себе нуждаются в содержательном обосновании; они лишь суммируют другой материал и дают начало логическому построению теории[2].

Необходимый период критики, систематизации и обоснования наступил для анализа к середине прошлого столетия. Усилиями ряда выдающихся ученых эта важная и трудная работа была успешно выполнена.

[1] Ф. Энгельс. Анти-Дюринг. Госполитиздат, 1953, стр. 34.

[2] Эту двоякую роль аксиом упускают иногда из вида даже в сочинениях методологического характера, придавая тем самым аксиоматическому построению несвойственное ему значение абсолютного обоснования теории.

В частности, получили строгие определения основные понятия действительного числа, переменной, функции, предела, непрерывности.

Впрочем, как мы уже имели случай заметить, никакое из этих определений нельзя считать абсолютно строгим и совершенно окончательным. Развитие этих понятий продолжается. Эвклид и все математики в течение двух тысяч лет после него, несомненно, считали эвклидовы «Начала» почти-что пределом логической строгости. Но теперь, на современный взгляд, эвклидово обоснование геометрии выглядит довольно поверхностным. Этот исторический пример учит, что не следует обольщаться на счет «абсолютной» и «окончательной» строгости современной математики. В науке, которая еще не умерла и не превратилась в мумию, нет и не может быть ничего вполне завершенного. Однако мы можем сказать с уверенностью, что, во-первых, установленные теперь основания анализа достаточно хорошо отвечают современным задачам науки и современному понятию о логической точности и что, во-вторых, продолжающееся углубление этих понятий и идущая вокруг них дискуссия не заставляют и не заставят просто отбросить эти основания; но они ведут к новому, более точному и глубокому их пониманию, о результатах которого в полной мере пока еще, может быть, трудно судить.

Хотя установление принципов теории — это итог ее развития, но он не служит ее концом, а напротив, служит новому ее движению. Так было с анализом. В связи с уточнением его основ возникла новая математическая теория — созданная немецким математиком Кантором в 70-х годах прошлого столетия общая теория бесконечных множеств любых абстрактных объектов, будь-то множества чисел, точек, функций или других «вещей» в том же роде. На почве этих идей выросла новая глава анализа — так называемая теория функций действительного переменного — понятие о которой, так же как понятие об основаниях анализа и теории множеств, изложено в главе XV (том 3). Вместе с тем общие идеи теории множеств проникли во все области математики. Но эта «теоретико-множественная точка зрения» неразрывно связана с новым этапом в развитии математики, к краткому рассмотрению которого мы сейчас и перейдем.

§ 7. СОВРЕМЕННАЯ МАТЕМАТИКА

1. Четырем этапам в развитии математики, о которых говорилось в § 5, естественно, отвечают ступени в математическом образовании, так что основному содержанию каждого из этих этапов можно довольно точно сопоставить уровень математических знаний, получаемых на разных ступенях обучения.

Основные результаты арифметики и геометрии, полученные в первый период развития математики, известны у нас всем и составляют предмет начального образования. Например, определяя количество материала, нужное для выполнения работы, скажем — настила пола, мы уже пользуемся этими первыми результатами математики:

Важнейшие достижения второго периода — периода элементарной математики — составляют предмет преподавания в средней школе.

Основные результаты третьего периода (основы анализа, теории дифференциальных уравнений, высшей алгебры и др.) составляют главное содержание математического образования каждого инженера; они изучаются так или иначе в каждой высшей школе, на каждом факультете, кроме гуманитарных. Таким образом, основные идеи и результаты математики этого периода широко известны, и ими пользуются в большем или меньшем объеме почти все инженеры и естествоиспытатели.

Напротив, идеи и результаты последнего, современного, этапа в развитии математики изучаются в основном только на специальных физико-математических факультетах. Кроме специалистов-математиков, ими пользуются научные работники в области механики, физики, ряде отраслей новой техники. Конечно, это вовсе не означает, что они удалены от приложений, но они представляют собой последние результаты развития науки и, естественно, оказываются более сложными. Поэтому, переходя сейчас к общей характеристике последнего этапа в развитии математики, мы не можем рассчитывать на то, что все, о чем мы коротко скажем, окажется вполне ясным. Мы попытаемся в нескольких штрихах дать только самую общую характеристику новых разделов математики, содержание которых более подробно выяснится из соответствующих глав книги.

Если этот параграф покажется излишне трудным, его можно в первом чтении пропустить, с тем чтобы вернуться к нему после ознакомления со специальными главами.

2. Начало современного этапа в развитии математики характеризовалось глубокими изменениями во всех ее основных разделах: алгебре, геометрии, анализе. Быть может наиболее отчетливо это изменение можно проследить на примере геометрии. В 1826 г. Лобачевским и почти одновременно также венгерским математикам Яношем Бойай была развита новая неэвклидова геометрия. Идеи Лобачевского далеко не сразу стали понятны всем математикам: они были слишком смелы и неожиданны. Однако именно с этого момента началось принципиально новое развитие геометрии, изменилось самое понимание того, что такое геометрия. Ее предмет и область применений стали быстро расширяться. Важнейший после Лобачевского шаг в этом направлении был сделан в 1854 г. знаменитым немецким математиком Риманом. Он явно формулировал общую идею о неограниченности числа «пространств», которые может изучать геометрия, и указал вместе с тем возможный их реальный смысл.

В новом развитии геометрии характерны два обстоятельства.

Во-первых, если прежде геометрия изучала только *пространственные* формы и отношения материального мира (притом лишь в той мере, в какой они отражаются в рамках эвклидовой геометрии), то теперь ее предмет составляют также многие *другие* формы и отношения действительности, лишь сходные с пространственными и потому допускающие использование при их исследовании геометрических методов. В связи с этим термин «пространство» приобрел в математике новый, более широкий и в то же время более специальный смысл. Одновременно сами методы геометрии стали во много раз богаче и разнообразнее. (В свою очередь они дают более совершенные средства познания того окружающего нас физического пространства, от которого была абстрагирована геометрия в ее первоначальном виде.)

Во-вторых, даже в эвклидовой геометрии произошли важные сдвиги: в ней изучаются свойства несравненно более сложных фигур, вплоть до произвольных точечных множеств. Появляется также принципиально новый подход к самим исследуемым свойствам фигур. Выделяются отдельные группы свойств, которые подвергаются исследованию в отвлечении от других, причем это отвлечение, это абстрагирование уже *внутри* геометрии порождает своеобразные ее разделы, являющиеся по существу самостоятельными «геометриями». Развитие геометрии во всех этих направлениях продолжается, и предметом ее рассмотрения служат все новые и новые «пространства» и их «геометрии»: пространство Лобачевского, проективное пространство, эвклидовы и другие пространства разных чисел измерения, например четырехмерное, римановы пространства, финслеровы пространства, топологические пространства и т. д. Эти теории находят важные применения как в самой математике, помимо геометрии, так и в физике и механике, причем особенно замечательным является их применение в теории относительности современной физической теории пространства, времени и тяготения. Из сказанного можно видеть, что речь идет о качественном изменении геометрии.

Идеи современной геометрии и некоторые элементы учения о разных исследуемых в ней пространствах будут изложены в главах XVII и XVIII (том 3).

3. Качественное изменение претерпела также алгебра. В первой половине прошлого столетия в ней зарождаются новые теории, которые привели к ее изменению, расширению ее предмета и области приложений.

Алгебра в своей первоначальной основе была, как уже говорилось в § 5, учением об арифметических действиях над числами, рассматриваемых формально, в общем виде, в отвлечении от данных конкретных чисел. Это отвлечение нашло отражение в том, что в алгебре величины обозначаются буквами, с которыми производят выкладки по известным формальным правилам.

Современная алгебра, сохраняя эту основу, колоссально расширяет ее. В ней рассматривают теперь «величины» гораздо более общей природы, чем числа, причем изучаются действия над этими «величинами», аналогичные в той или иной степени по своим формальным свойствам обычным арифметическим действиям сложения, вычитания, умножения и деления. Простейший пример представляют векторные величины, которые тоже, как известно, можно складывать по правилу параллелограмма. Но обобщение, проводимое в современной алгебре, таково, что даже самый термин «величина» часто теряет смысл, и говорят вообще об «объектах», над которыми можно производить действия, подобные обычным алгебраическим. Так, например, два движения, произведенные одно за другим, как очевидно, равносильны некоторому одному суммарному движению, два алгебраических преобразования формулы могут быть равносильны одному итоговому преобразованию и т. п. Соответственно можно говорить о своего рода «сложении» движений или преобразований. Все это и многое другое в том же роде изучает в общем отвлеченном виде современная алгебра.

Новые алгебраические теории, идущие в этом направлении, зародились в первой половине прошлого столетия в исследованиях ряда математиков, из которых особенно следует назвать французского математика Галуа (1811—1832). Понятия, методы и результаты современной алгебры находят существенные применения в анализе, геометрии, физике, кристаллографии и т. п. В частности, упомянутое в конце § 3 учение о симметрии кристаллов, развитое Е. С. Федоровым, опирается на соединение геометрии с одной из новых алгебраических теорий — так называемой теорией групп.

Мы видим, что речь идет о коренном, качественном обобщении предмета алгебры, об изменении самого понимания того, что такое алгебра. (Идеи современной алгебры и элементы некоторых ее теорий будут изложены в главах XX и XVI, том 3).

4. Анализ со всеми его ответвлениями также претерпел глубокие сдвиги. Во-первых, как уже было сказано в предыдущем параграфе, были уточнены его основания, в частности получили точные и общие определения его основные понятия: функция, предел, интеграл и, наконец, само понятие переменной величины (было дано строгое определение действительного числа). Начало уточнения основ анализа исходит от чешского математика Больцано (1781—1848), французского математика Коши (1789—1857) и ряда других. Это уточнение относится к тому же периоду, что и новое развитие алгебры и геометрии; оно было в известной мере завершено к 80-м годам прошлого столетия немецкими математиками Вейерштрассом, Дедекиндом и Кантором. Последний, как уже было сказано в конце § 6, положил начало теории бесконечных множеств, сыгравшей большую роль в развитии новых идей математики.

Уточнение понятий переменной и функции в связи с теорией множеств создало почву для дальнейшего развития анализа. Произошел переход к исследованию более общих функций; в соответствующем направлении обобщается аппарат анализа: интегральное и дифференциальное исчисления. Так, на пороге нашего столетия возникла уже упомянутая в § 6 новая глава анализа, называемая теорией функций действительного переменного. Развитие этой теории более всего обязано французским математикам Борелю, Лебегу и др. и потом особенно Н. Н. Лузину (1883—1950) и его школе. В целом новые главы анализа называют современным анализом, в отличие от прежнего, называемого классическим.

В анализе возникли и другие новые теории. Так, выделилась в особую область теория приближения функций, которая изучает вопросы о наилучшем приближенном представлении общих функций различными «простыми» функциями, в первую очередь многочленами, т. е. функциями вида

$$a_0x^n + a_1x^{n-1} + \ldots + a_{n-1}x + a_n.$$

Теория приближения функций имеет важное значение уже по одному тому, что дает общие основания для практического вычисления функций, для приближенной замены сложных функций функциями более простыми. Зачатки этой теории восходят еще ко времени возникновения анализа. Новое направление было дано ей великим русским математиком Пафнутием Львовичем Чебышевым (1821—1894). Направление это было позже развито в так называемую конструктивную теорию функцией, главным образом трудами советских математиков, особенно С. Н. Бернштейна (род. 1880), которому принадлежат здесь важнейшие результаты. Приближению функций посвящена глава XII (том 2).

Выше уже шла речь о развитии теории функций комплексного переменного. Мы должны еще упомянуть о так называемой качественной теории дифференциальных уравнений, берущей начало от работ Пуанкаре (1854—1912) и А. М. Ляпунова (1857—1918), о которой дается понятие в главе V (том 2), а также о теории интегральных уравнений и т. д. Последние теории имеют большое практическое значение для механики, физики и техники. Так, качественная теория дифференциальных уравнений решает задачи об устойчивости движения, работы механизмов, электроколебательных систем и т. п. Устойчивость процесса в самом общем смысле означает, что при малом изменении начальных данных или условий его течения весь его режим на всем протяжении времени будет изменяться также незначительно. Техническое значение такого рода вопросов едва ли нуждается в пояснениях.

5. На почве развития анализа и математической физики в соединении с новыми идеями геометрии и алгебры возникла новая обширная область — так называемый функциональный анализ, играющий исключительно важную роль в современной математике. В создании его принимали участие многие ученые; назовем, например, крупнейшего немецкого математика последнего времени Гильберта (1862—1943), венгерского математика

Рисса (1880—1956) и польского математика Банаха (1892—1945). Важные результаты в этой области, связанные с математической физикой, принадлежат молодым советским ученым. Функциональному анализу посвящена отдельная глава XIX (том 3).

Сущность этой новой части математики коротко состоит в следующем. Если в классическом анализе переменной является величина —«число», то в функциональном анализе уже самая функция рассматривается как переменная. Свойства данной функции определяются здесь не сами по себе, а по отношению этой функции к другим функциям. Поэтому рассматриваются уже не отдельные функции, а сразу совокупности всех функций, характеризуемых тем или иным общим свойством, например всех непрерывных функций. Такая совокупность функций объединяется в так называемое «функциональное пространство». Это соответствует, например, тому, что мы можем рассматривать совокупность всех кривых на плоскости или всех возможных движений данной механической системы, определяя свойства отдельных кривых, или движений по их отношениям к другим кривым, или другим движениям.

Переход от изучения или разыскания отдельных функций к рассмотрению *переменной* функции подобен переходу от неизвестных чисел x, y к *переменным* x, y, т. е. подобен указанной в предыдущем параграфе идее Декарта. В связи с этой идеей Декарт дал известное соединение алгебры и геометрии — уравнения с кривой, что послужило одним из решающих моментов возникновения анализа. Подобно этому теперь соединение понятия о переменной функции с идеями уже современной алгебры и современной геометрии породило новый, функциональный, анализ. Подобно тому, как анализ был необходим для развития создавшейся тогда механики, так функциональный анализ дал новые методы решения задач математической физики и представил математический аппарат для новой атомной, квантовой механики. История в известной мере повторяется, но по-новому, на более высокой ступени. Функциональный анализ соединяет, как мы уже сказали, основные идеи и методы анализа, современной алгебры и геометрии и в свою очередь оказывает влияние на их развитие. Задачи, идущие от классического анализа, получают теперь новые общие решения часто как раз посредством функционального анализа. Здесь, как в фокусе, собираются и дают практические плоды наиболее общие и абстрактные идеи современной математики.

Из этого краткого очерка, из одного перечисления новых направлений анализа (теория функций действительного переменного, теория приближения функций, качественная теория дифференциальных уравнений, теория интегральных уравнений, функциональный анализ) можно понять, что речь идет действительно о существенно новом этапе в развитии анализа.

6. Во все времена технический уровень средств вычисления оказывал существенное влияние на сами математические методы. Однако находящиеся в нашем распоряжении средства выполнения вычислений до последнего времени оставались весьма ограниченными. Простейшие приспособления (типа счетов), таблицы логарифмов и логарифмическая линейка, арифмометр, наконец, счетно-аналитические машины и автоматизированный арифмометр — вот основные средства вычислений, существовавшие к 40-м годам XX в. Эти средства обеспечивают более или менее быстрое выполнение отдельных операций (сложения, умножения и т. п.). Но доведение до численного результата практически возникающих задач требует подчас колоссального числа подобных операций, следующих сложной программе, зависящей иногда от получаемых по ходу дела результатов. Решение таких задач оказывалось практически недоступным или полностью обесценивалось длительностью самого решения.

В последнее десятилетие на наших глазах происходит коренное изменение всего уровня вычислительной техники. Современные вычислительные машины, построенные на новых принципах, позволяют вести вычисления с исключительно большой скоростью и при этом автоматически проводить сложные цепи вычислений по заранее намечаемым для этого весьма гибким программам. Некоторые вопросы, связанные с устройством и значением современных счетных машин, будут освещены в главе XIV.

Новая техника не только делает доступными неосуществимые ранее исследования, но она заставляет изменить оценку многих известных математических результатов. Ею особенно стимулируется развитие приближенных методов, т. е. путей, дающих возможность посредством цепи элементарных операций подойти к необходимому численному результату с достаточно большой точностью. Сами методы приходится при этом оценивать с точки зрения удобства их реализации на соответствующих машинах.

В тесной связи с развитием новой вычислительной техники находится математическая логика. Она развилась прежде всего из внутренних потребностей математики в связи с возникшими в ней трудностями и имеет своим предметом анализ математических доказательств. Будучи, собственно, разделом математики, математическая логика включает в себя те разделы общей логики, которые объективно допускают формализацию и могут развиваться математическим методом.

Восходя, с одной стороны, к истокам и основаниям математики, математическая логика, с другой стороны, оказывается тесно связанной с наиболее современными вопросами вычислительной техники. Естественно, например, что доказательство, ведущее к созданию определенного процесса, позволяющего приблизиться к результату с произвольно высокой точностью, существенно отличается от более отвлеченных доказательств существования того или иного результата.

Возникает также своеобразный круг вопросов в связи с исследованием пределов общности тех классов задач, которые вообще могут быть охвачены заведомо ведущим к результату единообразным, вполне определенным методом. На этом пути в математической логике получены глубокие результаты, весьма важные также с общей познавательной точки зрения.

Не будет преувеличением сказать, что в современной математике с развитием новой вычислительной техники и достижениями математической логики связан новый период, характеризующийся тем, что предметом исследования становится не только тот или иной объект, но и те пути, те формы, посредством которых этот объект задается, не только те или иные задачи, но и возможные способы их решения.

Ко всему сказанному остается только добавить, что и более старые области математики — теория чисел, эвклидова геометрия, классические алгебра и анализ, теория вероятностей — продолжают на протяжении всего периода современной математики бурно развиваться, обогащаясь новыми принципиальными идеями и результатами. Таковы, например, идеи и результаты, данные в теории чисел и наглядной геометрии русскими и советскими математиками П. Л. Чебышевым, Е. С. Федоровым, И. М. Виноградовым и другими. Широкое развитие теории вероятностей связано с чрезвычайно важными закономерностями статистической физики и современными техническими проблемами.

7. Каковы же наиболее общие характерные черты современной математики в целом, которые выявляются из только что рассмотренного развития геометрии, алгебры и анализа?

Это прежде всего громадное расширение предмета математики, расширение области ее приложений. Такое расширение предмета и области приложений математики означает вместе с тем громадный количественный и качественный ее рост, появление новых теорий и сильных математических методов, которые позволяют решать задачи, вовсе недоступные прежде. При этом расширение предмета математики характеризуется прежде всего тем, что современная математика сознательно ставит перед собой задачу изучения возможных типов количественных отношений и пространственных форм.

Другая характерная черта современной математики — это создание новых обобщающих понятий, новая, более высокая ступень абстракции. Именно эта особенность обеспечивает сохранение единства математики, несмотря на рост и разнообразие ее ответвлений. В областях, самых далеких друг от друга, обобщающие понятия и теории вскрывают единое и общее. Они же обеспечивают достаточную общность методов, широту приложений и глубокое взаимное проникновение основных разделов математики: геометрии, алгебры, анализа.

К характерным чертам современной математики следует отнести также известное господство теоретико-множественной точки зрения. Конечно, эта точка зрения сама приобретает смысл именно потому, что суммирует накопленный всем предшествующим развитием математик содержательный материал.

Наконец, одну из характерных черт современной математики состаı ляет более глубокий анализ ее основ, анализ взаимозависимости ее поня тий, структуры отдельных теорий, анализ самых способов математиче ских доказательств и выводов. Без такого анализа основ не могут совер шенствоваться и развиваться дальше сами обобщающие принципы и теории

Определяющую особенность современной математики можно видет в том, что ее предмет составляет уже не только данные, но и возможны количественные отношения и формы. В геометрии речь идет не тольк о пространственных, но и о сходных с пространственными, возможных отнс шениях и формах. В алгебре речь идет о разных системах абстрактны объектов с возможными законами действий над ними. В анализе перемен ной становится не только величина, но самая функция рассматривается как переменная. В функциональное пространство объединяются все функ ции того или иного типа, т. е. возможные зависимости между перемен ными. Поэтому коротко можно сказать, что если элементарная ма тематика есть математика постоянных величин, математика следующег периода — математика переменных величин, то *современная математика есть математика возможных, вообще говоря переменных, количественны отношений и взаимосвязей между величинами*. Это определение, конечно неполно, но оно в общем верно выделяет ту характерную черту совре менной математики, которая обусловливает ее качественное отличие от математики предыдущих эпох.

§ 8. СУЩНОСТЬ МАТЕМАТИКИ

1. Теперь мы можем, опираясь на все изложенное, перейти к общим выводам о сущности математики.

Сущность математики была выражена Энгельсом в одном из разделов «Анти-Дюринга», и мы приведем здесь этот замечательный отрывок.

В формулировках Энгельса читатель легко узнает то, что говорилось выше, например, по поводу арифметики и геометрии; и это понятно: мы излагали фактическую историю возникновения и развития математики, руководствуясь в ее понимании диалектическим материализмом. Диалектический материализм приводит к верным выводам именно потому, что он ничего не навязывает фактам, но рассматривает факты, как они есть, т. е. в их необходимых связях и развитии.

Свое изложение сущности математики Энгельс начинает с критических замечаний по поводу вздорных взглядов Дюринга, в частности по поводу ложного мнения, будто математика занимается творениями «чистого разума» независимо от опыта. Энгельс пишет:

«Но совершенно неверно, будто в чистой математике разум имеет дело только с продуктами собственного творчества и воображения. Понятия числа и фигуры взяты не откуда-нибудь, а только из действительного мира. Десять пальцев, на которых люди учились считать, т. е. производить первую арифметическую операцию, представляют собой все, что угодно, только не продукт свободного творчества разума. Чтобы считать, надо иметь не только предметы, подлежащие счету, но обладать уже способностью отвлекаться при рассматривании этих предметов от всех прочих их свойств кроме числа, а эта способность есть результат долгого, опирающегося на опыт, исторического развития. Как понятие числа, так и понятие фигуры заимствованы исключительно из внешнего мира, а не возникли в голове из чистого мышления. Должны были существовать вещи, имеющие определенную форму, и эти формы должны были подвергаться сравнению, прежде чем можно было дойти до понятия фигуры. Чистая математика имеет своим объектом пространственные формы и количественные отношения действительного мира, стало быть — весьма реальный материал. Тот факт, что этот материал принимает чрезвычайно абстрактную форму, может лишь слабо затушевать его происхождение из внешнего мира. Но чтобы быть в состоянии исследовать эти формы и отношения в чистом виде, необходимо совершенно отделить их от их содержания, оставить это последнее в стороне как нечто безразличное; таким путем мы получаем точки, лишенные измерений, линии, лишенные толщины и ширины, разные a и b, x и y, постоянные и переменные величины, и только в самом конце мы доходим до продуктов свободного творчества и воображения самого разума, а именно — до мнимых величин. Точно так же выведение математических величин друг из друга, кажущееся априорным, доказывает не их априорное происхождение, а только их рациональную взаимную связь. Прежде чем притти к мысли выводить *форму* цилиндра из вращений прямоугольника вокруг одной из его сторон, нужно было исследовать некоторое количество реальных прямоугольников и цилиндров, хотя бы и в очень несовершенных формах. Как и все другие науки, математика возникла из *практических нужд* людей: из измерения площадей земельных участков и вместимости сосудов, из счисления времени и из механики. Но, как и во всех других областях мышления, законы, абстрагированные от реального мира, на известной ступени развития отрываются от реального мира, противопоставляются ему как нечто самостоятельное, как явившиеся извне законы, с которыми мир должен сообразоваться. Так было с обществом и государством, так а не иначе, *чистая* математика *применяется* впоследствии к миру, хотя она заимствована из этого самого мира и только выражает часть присущих ему форм связей,— и собственно *только поэтому* может вообще применяться»[1].

2. Таким образом, Энгельс подчеркивает, что математика отражает действительность, что возникла она из практических нужд людей и воз-

[1] Ф. Энгельс. Анти-Дюринг, стр. 36—37.

никновение ее первых понятий и положений было результатом долгого, опирающегося на опыт исторического развития. Мы уже достаточно подробно проследили это на примере арифметики и геометрии.

Мы убедились, в частности, что именно так возникли понятия числа, величины, геометрической фигуры и что они отражают реальные количественные отношения и пространственные формы действительности. Точно так же основные понятия анализа отражают реальные количественные отношения, они складывались постепенно на основе обобщения громадного конкретного материала; так, понятие функции отражает в обобщенной, абстрактной форме резнообразные зависимости между реальными величинами.

Суммируя все это, Энгельс и приходит к основному выводу о том, что *математика имеет своим предметом определенный вполне реальный материал, но рассматривает его в полном отвлечении от конкретного содержания и качественных особенностей.* Этим, как мы видим, математика отличается от естественных наук, и Энгельс явно отделяет ее от естествознания [1].

Возможность такого абстрактного рассмотрения предмета математики имеет объективное основание в самом этом предмете. Те общие, не зависящие от качественных особенностей или конкретного содержания, формы, отношения, взаимосвязи и законы, которые отражаются в математике, существуют объективно, независимо от нашего сознания. Только существование числа как объективного свойства совокупности предметов, независимость взаимоотношений между числами от качественных особенностей предметов, богатство этих взаимоотношений сделали возможной арифметику. Там, где нет таких общих форм и отношений, безразличных к содержанию, невозможно и математическое рассмотрение.

3. Указанная основная особенность математики определяет другие характерные ее особенности. В § 2 мы рассматривали некоторые из них специально на примере арифметики. Это— специфический «язык формул», широта приложений, отвлеченный от опыта характер математических выводов, их логическая неизбежность и убедительность. Этот умозрительный характер математики является весьма существенной ее особенностью, и мы рассмотрим эту особенность подробнее.

Если мы отвлекли, скажем, понятие числа от его конкретных оснований и рассматриваем целые числа вообще, вне всякого отношения к тем или иным конкретным совокупностям предметов, то само собой ясно, мы не можем производить опытов над такими отвлеченными числами. Оставаясь на этом уровне абстракции и не возвращаясь к конкретным предметам, можно получать новые выводы о числах только путем рассуждения, исходя из самого *понятия* о числе. То же относится, конечно, ко всем другим математическим выводам. Оставаясь в пределах чистой геомет-

[1] См. Ф. Энгельс. Анти-Дюринг, стр. 10—11.

рии, т. е. рассматривая геометрические фигуры в полном отвлечении от всякого качественного, конкретного содержания, мы не можем получить новых выводов иначе как рассуждением, исходя из самого *понятия* о той или иной фигуре, из самих основных понятий или аксиом геометрии. Так, свойства круга выводят из понятия о круге как геометрическом месте точек, равноудаленных от данной точки, вовсе не думая уже о проверке каждой теоремы на опыте.

Стало быть, *отвлеченный характер математики уже предопределяет тот факт, что математические теоремы доказываются только рассуждением, исходя из самих понятий.*

Можно сказать, что в математике исследуют количественные отношения, имея в виду лишь то, что содержится в самом их определении. Соответственно математические выводы получают рассуждением, исходя из определений. Конечно, было бы неправильно понимать эти слова слишком буквально и предполагать, что достаточно строгие определения математических понятий действительно формулировались раньше, чем создавалась соответствующая математическая теория; на самом деле самые понятия уточнялись вместе с развитием теории, в результате ее развития. Глубокий анализ понятия о целом числе, так же как точная формулировка аксиом геометрии, были даны не в древности, а к концу XIX в. Тем более неверно думать, будто есть какое-либо абсолютно точно определенное математическое понятие. Всякое понятие, как бы ни казалось оно точно определенным, все-таки подвижно, оно развивается и уточняется с развитием науки. Это вполне доказано развитием математики в отношении всех ее понятий и это только лишний раз подтверждает основное положение диалектики о том, что нет на свете ничего такого, что было бы совершенно неподвижно и никак не развивалось бы. Поэтому и в отношении математических понятий можно говорить, во-первых, только о достаточной, но никак не совершенной их определенности, а во-вторых, нужно иметь в виду, что точность и ясность их определения, глубина их анализа развиваются с развитием математики. На этой подвижности математических понятий мы еще будем иметь случай остановиться в следующем параграфе, а сейчас, имея в виду сделанное замечание, обращаем внимание именно на достаточную их определенность.

Именно эта определенность математических понятий вместе с общезначимостью самой логики оказываются причиной характерной для математики внутренней убедительности и логической необходимости ее выводов. Неизбежность умозрительных выводов математики дает повод к ошибочному представлению, будто математика имеет основание в чистом мышлении, будто она априорна, а не исходит из опыта, будто она не отражает действительности. К такого рода взглядам пришел, например, знаменитый немецкий философ Кант. Это глубоко ошибочное идеалистическое представление происходит, в частности, от того, что математику рассматривают не в ее реальном возникновении и развитии, а в готовом виде. Но такой подход совершенно несостоятелен уже по той простой причине, что не соответствует фактическому положению дел. То, что математика не априорна, а возникла из опыта, — это твердо установленный факт. Кстати, о фактическом возникновении геометрии писал еще Эвдем Родосский, которого мы цитировали в § 3.

Не только самые понятия математики, но и ее выводы, ее методы отражают действительность. Это важное обстоятельство как раз и вскрывает Энгельс, когда пишет, что «выведение математических величин друг из друга, кажущееся априорным, доказывает не их априорное происхождение, а только их рациональную взаимную связь». Математические выводы и доказательства возникли как отражение реальных связей, которые люди исследовали на опыте. Сложение чисел отражает реальное соединение нескольких совокупностей предметов в одну. Известное доказательство теоремы о равенстве треугольников, в котором говорят об их наложении, несомненно, имеет своим источником операцию фактического прикладывания предметов друг к другу, которая постоянно производится при сравнении их размеров. Вычисление объемов интегрированием отражает в абстрактной форме реальную возможность складывать тела из тонких слоев или резать их на такие слои. Более сложные математические доказательства есть результат дальнейшего развития, исходящего из таких материальных оснований.

4. Полное отвлечение предмета математики от всякой конкретности и основанный на этом умозрительный характер математических выводов влекут за собою другую важную особенность математики: в математике исследуют не только такие количественные отношения и пространственные формы, которые непосредственно абстрагируются из действительности, но и такие отношения и формы, которые определяются внутри самой математики на основе уже сложившихся математических понятий и теорий. Именно на эту особенность математики обращает внимание Энгельс, когда, указав на возникновение понятий точки, линии, постоянной и переменной величины, говорит: «... и только в самом конце мы доходим до продуктов свободного творчества и воображения самого разума, а именно — до мнимых величин».

Историческим фактом является то, что мнимые числа не были взяты из действительности в том же смысле, как, скажем, целые числа. Они появились первоначально внутри самой математики, из необходимого развития алгебры, как корни уравнений вида $x^2 = -a$ (где $a > 0$). И хотя постепенно с ними начали оперировать довольно свободно, их реальный смысл оставался долго неясным, почему за ними и закрепилось название «мнимых». Потом было открыто их геометрическое истолкование и они нашли многочисленные важные применения. Точно так же геометрия Лобачевского возникла как продукт творчества этого великого ученого; он не видел еще ее реального значения и называл ее потому «воображаемой геометрией». Но она была не свободной игрой ума, а неизбежным выводом из основных понятий геометрии, и Лобачевский рассматривал ее как возможную теорию пространственных форм и отношений. Поэтому «свободное творчество и воображение», о которых говорит Энгельс, нельзя понимать как простой произвол мысли. Свободное творчество в науке — это осознанная логическая необходимость, определяющаяся исходными, взятыми из опыта понятиями и положениями.

На новом этапе развития математики, начало которому положило как раз построение геометрии Лобачевского и точной теории мнимых чисел, возникли и постоянно возникают новые понятия и теории, создаваемые на основе уже сложившихся понятий и теорий без того, чтобы заимствовать их непосредственно из действительности. Математика определяет и исследует возможные формы действительности, что как раз и составляет одну из решающих особенностей последнего этапа её развития.

Правильное понимание этой особенности дает теория познания диалектического материализма. Ленин писал: «Познание есть отражение человеком природы. Но это не простое, не непосредственное, не цельное отражение, а процесс ряда абстракций, формирования, образования понятий, законов....» [1]. Метафизический материализм также признает познание, в частности математику, отражением природы. Однако, как отмечает Ленин, беда метафизического материализма состоит в неумении применить диалектику к теории отражения [2]. Метафизический материализм не понимает сложности этого отражения, не понимает того, что оно идет через ряд абстракций, путем формирования новых понятий, построения новых теорий на основе уже сложившихся понятий и теорий, путем рассмотрения не только данного в опыте, но и возможного. Между тем такой переход от данного к возможному обнаруживается уже в образовании таких понятий, как любое целое число или бесконечная прямая, потому что в опыте не даны ни сколь угодно большие числа, ни бесконечные прямые. Но когда понятие числа выкристаллизовалось, то из самого этого понятия, из самого закона образования последовательных чисел путем прибавления единицы выявилась возможность бесконечного продолжения числового ряда. Совершенно так же из проведения прямых выявилась возможность неограниченного продолжения прямой, выраженная во втором постулате Эвклида: «каждую прямую можно неограниченно продолжить». Дальнейший процесс абстракции привел к понятиям о *всем* натуральном ряде чисел и о *всей* бесконечной прямой. На последнем этапе развития математики качественно новым явилось построение теорий, идущих через ряд абстракций и формирования понятий. Но, восходя по этим ступеням абстракции, математика вовсе не отрывается от действи-

[1] В. И. Ленин. Философские тетради, стр. 156.

[2] Там же, стр. 330.

тельности. Новое вырастает в ней на основе отражения действительности, вследствие логики самого ее предмета, и именно в силу этого возвращается к действительности в применениях к проблемам физики и техники. Так было с мнимыми числами. То же верно в отношении других математических теорий, как бы ни были они абстрактны.

Характерный пример представляют теории различных многомерных пространств. Они складывались как обобщение эвклидовой геометрии, в соединении с развитием алгебры и анализа, под влиянием механики и физики. Сочетание этих идей привело Римана к построению общей теории, которая была развита дальше другими математиками, нашла ряд важных приложений и, наконец, послужила готовым математическим аппаратом для построения Эйнштейном общей теории относительности, точнее, теории тяготения. Абстрактные геометрические теории нашли такие блестящие приложения не случайно, не вследствие «предустановленной гармонии природы и разума», а вследствие того, что сами они выросли на почве геометрии, возникшей непосредственно из опыта, и в своем возникновении связывались их творцами с задачей исследования реального пространства. Риман, в частности, прямо предвидел связь своей теории с теорией тяготения.

Так, в развитии математики осуществляется закон движения познания, формулированный В. И. Лениным: «Мышление, восходя от конкретного к абстрактному, не отходит — если оно *правильное* ... — *от* истины, а подходит к ней. Абстракция *материи*, *закона* природы, абстракция *стоимости* и т. д., одним словом *все* научные (правильные, серьезные, не вздорные) абстракции отражают природу глубже, вернее, *полнее*. От живого созерцания к абстрактному мышлению *и от него к практике* — таков диалектический путь познания *истины*, познания объективной реальности»[1].

Из сказанного ясно, что совершенно ложным является идеалистический взгляд, будто математические теории представляют собой только условные схемы, предназначенные для описания данных опыта или «упорядочения потока ощущений» на основе «принципа экономии мышления».

Энгельс отмечает (см. цитату на стр. 61), что положения математики, абстрагированные от реального мира, как бы противопоставляются ему и применяются к его изучению, как некоторые готовые схемы. Мы, действительно, постоянно пользуемся, например, счетом, применяя его в готовом виде. Тем более это верно в отношении теорий, возникающих на более высоких ступенях абстракции. В качестве примера уже упоминалось, что риманова геометрия послужила *готовой* математической схемой для теории тяготения. Но Энгельс объясняет, что возможность такого применения математики к исследованию реального мира основана на том, что она заимствована из этого самого мира и только выражает часть при-

---

[1] В. И. Ленин. Философские тетради, стр. 146—147.

сущих ему форм связей и собственно *только потому* может вообще применяться. Тот факт, что многие теории создаются внутри самой математики, ничего здесь не меняет. Возникая как теории возможных форм действительности, они вовсе не условны, потому что возникают необходимо, вследствие самой логики предмета и именно поэтому находят реальные применения. Так или иначе, математические теории отражают действительность и различие состоит лишь в том, что это отражение в одних случаях более непосредственно, тогда как в других идет через ряд абстракций, образования понятий и т. д.

5. Последний этап в развитии математики характерен не только более высокими ступенями абстракции, он характерен еще существенным расширением ее предмета, выходящего за рамки первоначального понимания количественных отношений и пространственных форм.

Фигуры в многомерных или бесконечномерных пространствах — это, конечно, не пространственные формы в обычном смысле, как их понимаем мы все, когда имеем в виду обычное реальное пространство, а не абстрактные пространства математики. Эти пространства имеют реальный смысл и отражают в отвлеченном виде определенные формы действительности, но эти формы только сходны с пространственными; поэтому в отношении к обычному реальному пространству их можно назвать «пространственно-подобными». Говоря о многомерном пространстве и о фигурах в нем, мы тем самым придаем понятию пространства новое содержание, так что необходимо ясно различать обобщенное абстрактное понятие пространства в математике, с одной стороны, и понятие пространства в его исходном смысле универсальной формы существования материи, с другой.

Другим примером выхода предмета математики за пределы пространственных форм и количественных отношений в первоначальном смысле этих слов может служить возникновение в конце прошлого века новой дисциплины — математической логики, достигшей теперь широкого развития. Предметом ее рассмотрения является строение математических выводов, иными словами, она изучает, какие предложения можно выводить из данных посылок данными средствами. Она исследует свой предмет, как это свойственно именно математике, в полном отвлечении от содержания и потому заменяет предложения формулами, а правила умозаключения — правилами оперирования с этими формулами. Отношения между посылками и заключением, аксиомами и теоремами, конечно, не сводятся к пространственным формам или в обычном смысле к количественным отношениям, скажем, к отношениям объемов понятий.

В качестве другого примера укажем на теорию групп, которую можно понимать как учение о симметрии в самом общем виде. Однако изменение симметрии кристалла, скажем, при переходе серы из ромбической формы в призматическую, есть коренное качественное изменение состояния вещества. Таким образом, теория групп есть учение о таких величинах или о таких определенностях предметов, изменение которых сопровождается коренным изменением самих предметов.

Итак, расширение предмета математики ведет к существенному расширению самого понятия количественных отношений и пространственных форм. Каковы же в таком случае характерные общие черты этого расширяющегося предмета математики?

Если отвечать на этот вопрос не перечислением, а постараться выяснить то общее и характерное, что есть в предмете математики при всем его разнообразии, то ответ мы находим по существу у Энгельса. Достаточно принять во внимание не только его указание на предмет математики, но также и на способ рассмотрения этого предмета: полное отвлечение форм и отношений от содержания. Этот абстрактный характер математики дает одновременно также определение ее предмета.

Предмет математики составляют те формы и отношения действительности, которые объективно обладают такой степенью безразличия к содержанию, что могут быть от него полностью отвлечены и определены в общем виде с такой ясностью и точностью, с сохранением такого богатства связей, чтобы служить основанием для чисто логического развития теории. Если такие отношения и формы и называть количественными в общем смысле слова, то можно коротко сказать, что математика имеет своим предметом количественные отношения и формы, взятые в их чистом виде.

Абстракция отнюдь не является привилегией математики. Однако другие науки интересуются прежде всего соответствием своих абстрактных схем какому-либо вполне определенному кругу явлений и включают как одну из важнейших задач исследование границ применимости к данному кругу явлений уже сложившейся системы понятий и соответствующей смены применяемой системы абстракций. Математика, напротив, исследуя общие свойства в полном отвлечении от конкретных явлений, рассматривает сами эти системы абстракций в их отвлеченной общности, вне границ их применимости к отдельным конкретным явлениям. Можно сказать, что для математики характерно своего рода абсолютизирование ее абстракций.

Именно указанное объективное безразличие к содержанию исследуемых в математике форм определяет основные особенности математики: ее умозрительный характер, логическую необходимость и кажущуюся непреложность ее выводов, возникновение внутри нее новых понятий и теорий; этим же безразличием к содержанию обусловлены особенности приложений математики. Когда мы смогли перевести практическую задачу на язык математики, мы одновременно смогли отвлечься от второстепенных конкретных особенностей задачи и, пользуясь общими формулами и выводами, получить определенный результат. Отвлеченность математики составляет, таким образом, ее силу, и эта отвлеченность практически необходима.

6. Возвращаясь теперь к суждению Энгельса о математике, мы видим, какая глубина и богатство содержания, какие возможности развития заключаются в этом суждении. Не будучи сам математиком, он дал столь глубокий анализ основ этой науки не только потому, что был гениальным мыслителем, но, что самое главное, потому, что владел диалектическим материализмом и руководствовался им в задаче выяснения сущности математики. Не мудрено поэтому, что никто до Энгельса и не мог дать столь глубокого и верного решения этого вопроса. Самые крупные математики не могли его решить в таком объеме.

Точно так же в дальнейшем Ленин дал такой анализ проблемы физики, который превосходит все, сделанное в этой области.

Это доказывает лишний раз значение и силу диалектического материализма; это доказывает, что для овладения наукой недостаточно знания ее отдельных положений, недостаточно даже быть творческим работником в этой науке — для этого нужно еще владеть верным общим методом, владеть диалектическим материализмом. Без этого выводы науки либо будут казаться бесформенной грудой, либо представляться в искаженном виде; вместо верного понимания науки получится ложное, метафизическое, идеалистическое представление о ней. Так, например, многие математики, не владеющие диалектическим материализмом, либо вовсе не ориентируются в общих вопросах своей науки, либо трактуют их совершенно неверно [1].

В то время, когда Энгельс писал «Анти-Дюринг», т. е. в 1876—1877 гг., неэвклидова геометрия и геометрия многомерных пространств только что получили признание среди математиков, теория групп только оформилась, теория множеств только что возникала, а математическая логика лишь зарождалась. Поэтому понятно, что особенности нового этапа в развитии математики не могли быть детально отражены Энгельсом; и тем не менее, в его суждениях мы находим указания и для их понимания.

§ 9. ЗАКОНОМЕРНОСТИ РАЗВИТИЯ МАТЕМАТИКИ

В заключение мы попытаемся в кратких чертах охарактеризовать общие закономерности развития математики.

1. Математика не есть создание какой-либо одной исторической эпохи, какого-либо одного народа; она есть продукт ряда эпох, продукт работы многих поколений. Ее первые понятия и положения возникли,

---

[1] Любопытно, например, отметить, что два известных американских геометра Веблен и Уайтхед в своей книге «Основания дифференциальной геометрии» пытаются подойти к определению того, что такое геометрия, и приходят к выводу, что такого определения дать нельзя, кроме разве следующего: «геометрия есть то, что называют геометрией специалисты».

как мы видели, в глубокой древности и уже более двух тысяч лет назад были приведены в стройную систему. Несмотря на все преобразования математики, ее понятия и выводы сохраняются, переходя из одной эпохи к другой, как, например, правила арифметики или теорема Пифагора.

Новые теории включают в себя предшествующие достижения, уточняя, дополняя и обобщая их.

В то же время, как ясно из данного выше краткого очерка истории математики, ее развитие не только не сводится к простому накоплению новых теорем, но включает существенные, качественные изменения. Соответственно, развитие математики разделяется на ряд периодов, переходы между которыми как раз и обозначены такими коренными изменениями в самом предмете или структуре этой науки.

Математика включает в свою сферу все новые области количественных отношений действительности. В то же время важнейшим предметом математики были и остаются пространственные формы и количественные отношения в простом, наиболее непосредственном смысле этих слов, и математическое осмысление новых связей и отношений неминуемо происходит на основе и в связи с уже сложившейся системой количественных и пространственных научных представлений.

Наконец, накопление результатов внутри самой математики необходимо влечет как восхождение к новым ступеням абстракции, к новым обобщающим понятиям, так и углубление в анализ основ и первоначальных понятий.

Как дуб в своем могучем росте утолщает старые ветви новыми слоями, выбрасывает новые ветви, тянется вверх и углубляется корнями вниз, так и математика в своем развитии накапливает новый материал в уже сложившихся своих областях, образует новые направления, восходит к новым вершинам абстракции и углубляется в своих основах.

2. Математика имеет своим предметом реальные формы и отношения действительности, но, как говорил Энгельс, чтобы изучить эти формы и отношения в чистом виде, необходимо совершенно отделить их от их содержания, оставить это последнее в стороне как нечто безразличное. Однако форм и отношений вне содержания не существует, математические формы и отношения не могут быть абсолютно безразличными к содержанию. Стало быть, математика, по самой своей сущности стремящаяся осуществить такое отделение, стремится осуществить невозможное. *Это и есть коренное противоречие в самой сущности математики.* Оно является специфическим для математики проявлением общего противоречия познания. Отображение мыслью всякого явления, всякой стороны, всякого момента действительности огрубляет, упрощает его, выхватывая его из общей связи природы. Когда люди, изучая свойства пространства, установили, что оно имеет эвклидову геометрию, был совершен исклю-

чительно важный акт познания, но в нем же заключалось заблуждение: реальные свойства пространства были взяты упрощенно, схематично, в отвлечении от материи. Но без этого просто не было бы геометрии, и именно на почве этого отвлечения (как из внутреннего его исследования, так и из сопоставления математических результатов с новыми данными других наук) зарождались и укреплялись новые геометрические теории.

Постоянное разрешение и восстановление указанного противоречия на все более приближающихся к действительности ступенях познания и составляет сущность развития познания. При этом определяющим является, конечно, положительное содержание познания, элемент абсолютной истины в нем. Познание идет по восходящей линии, а не топчется на месте в простом смешении с заблуждением. Движение познания есть постоянное преодоление его неточности и ограниченности.

Указанное основное противоречие влечет за собой другие. Мы видели это на примере противоположностей дискретного и непрерывного. (В природе между ними нет абсолютного разрыва, и их разделение в математике неизбежно влекло необходимость создания все новых понятий, глубже отражающих действительность и одновременно преодолевающих внутренние несовершенства существующей математической теории). Совершенно так же противоречия конечного и бесконечного, абстрактного и конкретного, формы и содержания и др. выступают в математике как проявления ее коренного противоречия. Но решающее его проявление состоит в том, что, отвлекаясь от конкретного, вращаясь в кругу своих абстрактных понятий, математика тем самым отделяется от эксперимента и практики, а вместе с тем она лишь постольку является наукой (т. е. имеет познавательную ценность), поскольку опирается на практику, поскольку оказывается не чистой, а прикладной математикой. Говоря несколько гегелевским языком, *чистая математика постоянно «отрицает» себя как чистую математику; без этого она не может иметь научного значения, не может развиваться, не может преодолевать неминуемо возникающие внутри нее трудности.*

В своем формальном виде математические теории противостоят реальному содержанию как некоторые схемы для конкретных выводов. Математика выступает при этом как метод формулировки количественных законов естествознания, как аппарат для разработки его теорий, как средство решения задач естествознания и техники. Значение чистой математики на современном этапе заключено прежде всего в математическом методе. И как всякий метод существует и развивается не сам по себе, а только на основе своих применений, в связи с содержанием, к которому он применяется, так и математика не может существовать и развиваться без применений. Здесь опять обнаруживается единство противоположностей: общий метод противостоит конкретной задаче, как средство ее решения, но он сам возникает из обобщения конкретного материала и существует, развивается и находит свое оправдание только в решении конкретных задач.

3. Общественная практика играет определяющую роль в развитии математики в трех отношениях. Она ставит перед математикой новые проблемы, стимулирует ее развитие в том или ином направлении и дает критерий истинности ее выводов.

Это чрезвычайно ясно видно на примере возникновения анализа. Во-первых, именно развитие механики и техники выдвинуло проблему изучения зависимостей переменных величин в их общем виде. Архимед, подойдя вплотную к дифференциальному и интегральному исчислению, оставался, однако, в рамках задач статики, тогда как в новое время именно исследование движения породило понятия переменной и функции и понудило к оформлению анализа. Ньютон не мог развить механику, не развивая соответствующего математического метода.

Во-вторых, именно потребности общественного производства побуждали к постановке и решению всех этих проблем. Ни в античном, ни в средневековом обществе этих стимулов еще не было. Наконец, весьма характерно, что математический анализ в своем возникновении находил обоснование своих выводов именно в приложениях. Только поэтому он и мог развиваться без тех строгих определений его основных понятий (переменная, функция, предел), которые были даны позже. Истинность анализа устанавливалась применениями в механике, физике и технике.

Сказанное относится ко всем периодам развития математики. Начиная с XVII в. наиболее непосредственное влияние на ее развитие оказывают вместе с механикой теоретическая физика и проблемы новой техники. Механика сплошной среды, а потом теория поля (теплопроводность, электричество, магнетизм, поле тяготения) направляют развитие теории дифференциальных уравнений в частных производных. Разработка молекулярной теории и вообще статистической физики, начиная с конца прошлого века, служила важным стимулом развития теории вероятностей, особенно теории случайных процессов. Теория относительности сыграла решающую роль в развитии римановой геометрии с ее аналитическими методами и обобщениями.

В настоящее время развитие новых математических теорий, как функциональный анализ и др., стимулируется проблемами квантовой механики и электродинамики, задачами вычислительной техники, статистическими вопросами физики и техники и т. д. и т. п. Физика и техника не только ставят перед математикой новые задачи, наталкивают ее на новые предметы исследования, но также пробуждают развитие нужных для них разделов математики, которые складывались первоначально в большей мере внутри нее самой, как это было с римановой геометрией. Короче, для интенсивного развития науки нужно, чтобы она не только подошла к решению новых задач, но чтобы необходимость их решения навязывалась потребностями развития общества. В математике в последнее время возникает много теорий, но только те из них получают развитие и прочно входят в науку, которые нашли свои применения в естествознании и технике либо сыграли роль важных обобщений тех теорий, которые имеют такие приложения. Вместе с тем другие теории остаются без движения, как, например, некоторые рафинированные геометрические теории (недезарговы, неархимедовы геометрии), не нашедшие существенных применений.

Истинность математических выводов находит свое последнее основание не в общих определениях и аксиомах, не в формальной строгости доказательств, а в реальных приложениях, т. е. в конечном счете в практике.

В целом, развитие математики нужно понимать прежде всего как результат взаимодействия логики ее предмета, отраженной во внутренней логике самой математики, влияния производства и связей с естествознанием. Это различие идет сложными путями борьбы противоположностей, включая существенные изменения в основном содержании и формах математики. По содержанию развитие математики определяется ее предметом, но побуждается оно в основном и в конечном счете потребностями производства. Такова основная закономерность развития математики.

Конечно, мы не должны забывать при этом, что речь идет лишь об основной закономерности и что связь математики с производством, вообще говоря, является сложной. Из того, что говорилось выше, ясно, что было бы наивным пытаться обосновать появление каждой данной математической теории непосредственным «производственным заказом». Более того, математика, как и всякая наука, обладает относительной самостоятельностью, своей внутренней логикой, отражающей, как мы это подчеркивали, объективную логику, т. е. закономерность ее предмета.

4. Математика всегда испытывала самое существенное влияние не только общественного производства, но и всех общественных условий в целом. Ее блестящий прогресс в эпоху возвышения древней Греции, успехи алгебры в Италии в эпоху Возрождения, развитие анализа в эпоху, последовавшую за английской революцией, успехи математики во Франции в период, примыкающий к Французской революции, — все это убедительно демонстрирует неразрывную связь прогресса математики с общим техническим, культурным, политическим прогрессом общества.

Это также ярко видно на примере развития математики в России. Становление самостоятельной русской математической школы, идущей от Лобачевского, Остроградского и Чебышева, нельзя отделить от прогресса русского общества в целом. Время Лобачевского — это время Пушкина, Глинки, время декабристов, и расцвет математики был одним из элементов общего подъема.

Тем более убедительно влияние общественного развития в период после Великой Октябрьской социалистической революции, когда исследования фундаментального значения появлялись друг за другом с поразительной быстротой во многих направлениях: в теории множеств, топологии, теории чисел, теории вероятностей, теории дифференциальных уравнений, функциональном анализе, алгебре, геометрии.

Наконец, математика всегда испытывала и испытывает на себе заметное влияние идеологии. Как и во всякой науке, объективное содержание математики воспринимается и толкуется математиками и философами в рамках той или иной идеологии.

Короче, объективное содержание науки всегда укладывается в те или иные идеологические формы; единство и борьба этих диалектических противоположностей — объективного содержания и идеологических форм — в математике, как и во всякой науке, играют далеко не последнюю роль в ее развитии.

Борьба материализма, отвечающего объективному содержанию науки, с идеализмом, противоречащим этому содержанию и извращающим его понимание, идет через всю историю математики. Эта борьба ясно обозначена уже в древней Греции, где против материализма Фалеса, Демокрита и других философов, создававших греческую математику, выступал идеализм Пифагора, Сократа и Платона. С развитием рабовладельческого строя верхушка общества отрывалась от участия в производстве, считая его уделом низшего класса, и это порождало отрыв «чистой» науки от практики. Достойной внимания истинного философа признавалась лишь чисто теоретическая геометрия. Характерно, что появившиеся исследования некоторых механических кривых и даже конических сечений Платон считал остающимися за пределами геометрии, так как они «не приводят нас в общение с вечными и бестелесными идеями» и «нуждаются в применении орудий пошлого ремесла».

Яркий пример борьбы материализма против идеализма в математике представляет деятельность Лобачевского, который выдвинул и отстаивал материалистическое понимание математики против идеалистических взглядов кантианства.

Для русской математической школы вообще характерна материалистическая традиция. Так, Чебышев явно подчеркивал решающее значение практики, а Ляпунов выразил стиль отечественной математической школы в следующих замечательных словах: «Детальная разработка вопросов, особенно важных с точки зрения приложения и в то же время представляющих особенные теоретические трудности, требующие изобретения новых методов и восхождения к принципам науки, затем обобщение полученных выводов и создание этим путем более или менее общей теории». Обобщения и абстракции не сами по себе, а в связи с конкретным материалом, теоремы и теории не сами по себе, а в общей связи науки, ведущей в конечном счете к практике,— вот что оказывается на самом деле важным и перспективным [1].

Таковы же были устремления таких великих ученых, как Гаусс и Риман.

Однако с развитием капитализма в Европе материалистические взгляды, отражавшие передовую идеологию возвышающейся буржуазии эпохи XVI — начала XIX вв., стали сменяться идеалистическими воззрениями. Так, например, Кантор (1846—1918), создавая теорию бесконечных множеств, прямо ссылался на бога, высказываясь в том духе, что бесконечные множества имеют абсолютное существование в божественном разуме. Крупнейший французский математик конца XIX— начала XX в. Пуанкаре выдвинул идеалистическую концепцию «конвенционализма», согласно которой математика есть схема условных соглашений, принимаемых для удобства описания многообразия опыта. Так, по мнению Пуанкаре, аксиомы эвклидовой геометрии суть не более как условные соглашения и значение их определяется удобством и простотой, но не соответствием реальной действительности. Поэтому Пуанкаре говорил, что, например, в физике скорее откажутся от закона прямолинейного распространения света, чем от эвклидовой геометрии. Эта точка зрения была опровергнута развитием теории относительности, которая, вопреки всей «простоте» и «удобству» эвклидовой геометрии, в полном согласии с материалистическими идеями Лобачевского и Римана, привела к выводу, что реальная геометрия пространства отлична от эвклидовой.

На почве трудностей, возникших в теории множеств, и в связи с необходимостью анализа основных понятий математики, среди математиков в начале XX в. появились разные течения. Единство в понимании содержания математики было утрачено; разные математики стали по-разному рассматривать не только общие основы науки, что было и раньше, но даже по-разному стали оценивать смысл и значение отдельных конкретных результатов и доказательств. Выводы, казавшиеся осмысленными и содержательными для одних, другие объявляли лишенными смысла и значения. Возникли идеалистические течения «логицизма», «интуиционизма» «формализма» и др.

Логисты утверждают, что вся математика выводима из понятий логики. Интуиционисты видят источник математики в интуиции и придают смысл лишь интуитивно воспринимаемому. Поэтому они, в частности, вовсе отрицают значение канторовской теории бесконечных множеств. Более того, интуиционисты отрицают простой смысл даже таких утверж-

---

[1] Вообще понимание необходимых связей отдельных областей математики друг с другом, с естествознанием и практикой имеет чрезвычайно большое значение не только для верного взгляда на саму математику, но и для ориентировки ученых в выборе направлений и предметов исследования.

дений, как теорема о том, что всякое алгебраическое уравнение n-й степени имеет n корней. Для них это утверждение пусто, пока не указан способ вычисления корней. Так, полное отрицание объективного смысла математики привело интуиционистов к опорочиванию, как «лишенной смысла», значительной части достижений математики. Наиболее крайние из них дошли до утверждения, что существует столько математик, сколько есть математиков.

Попытку по-своему спасти математику от такого рода нападок предпринял крупнейший математик начала нашего века — Д. Гильберт. Сущность его идеи сводилась к тому, чтобы свести математические теории к чисто формальным операциям над символами согласно предписанным правилам. Расчет состоял в том, что при таком совершенно формальном подходе все трудности будут сняты, ибо предметом математики окажутся символы и правила действия с ними без всякого отношения к их смыслу. Это и есть установка формализма в математике. По словам интуициониста Брауэра, для формалиста истина математики на бумаге, тогда как для интуициониста она в голове математика.

Нетрудно, впрочем, видеть, что оба они неправы, ибо математика, а вместе с тем и то, что написано на бумаге, и то, что думает математик, отражает действительность, и истина математики заключается в ее соответствии объективной действительности. Отрывая математику от материальной действительности, все эти течения оказываются идеалистическими.

Идея Гильберта потерпела поражение в результате ее собственного развития. Австрийский математик Гедель доказал, что даже арифметику нельзя формализовать полностью, как на то рассчитывал Гильберт. Вывод Геделя явно вскрыл внутреннюю диалектику математики, которая не позволяет исчерпать ни одну ее область формальным исчислением. Даже простейшая бесконечность натурального ряда чисел оказалась неисчерпываемой конечной схемой символов и правил действия с ними. Так, было *математически* доказано то, что высказал в общем виде еще Энгельс, когда писал:

«Бесконечность *есть* противоречие... Уничтожение этого противоречия было бы концом бесконечности»[1]. Гильберт рассчитывал заключить математическую бесконечность в рамки конечных схем и тем самым ликвидировать все противоречия и трудности. Это оказалось невозможным.

Но в условиях капитализма конвенционализм, интуиционизм, формализм и другие подобные течения не только сохраняются, но и дополняются новыми вариантами идеалистических взглядов на математику. Теории, связанные с логическим анализом основ математики, существенно используются в некоторых новых вариантах субъективного идеализма. Субъек-

[1] Ф. Энгельс. Анти-Дюринг, стр. 49.

тивный идеализм использует теперь математику, в частности математическую логику, не меньше, чем физику, и потому вопросы понимания основ математики приобретают особую остроту.

Так, трудности развития математики породили в условиях капитализма идеологический кризис этой науки, сходный в своих основах с кризисом физики, сущность которого была выяснена Лениным в его гениальном произведении «Материализм и эмпириокритицизм». Этот кризис вовсе не означает, что математика в капиталистических странах совершенно задержана в своем развитии. Ряд ученых, стоящих на явно идеалистических позициях, делает важные, порой выдающиеся успехи в решении конкретных математических вопросов и развитии новых теорий. Достаточно сослаться на блестящую разработку математической логики.

Коренной порок распространенного в капиталистических странах взгляда на математику состоит в его идеализме и метафизике: в отрыве математики от действительности и пренебрежении ее реальным развитием. Логистика, интуиционизм, формализм и другие подобные направления выделяют в математике какую-нибудь одну ее сторону — связь с логикой, интуитивную ясность, формальную строгость и т. п.,— неосновательно преувеличивают, абсолютизируют ее значение, отрывают ее от действительности и за глубоким анализом этой одной черты математики самой по себе теряют из виду математику в целом. Именно вследствие этой односторонности ни одно из этих течений при всей тонкости и глубине отдельных выводов не может привести к верному пониманию математики. В противоположность различным течениям и оттенкам идеализма и метафизики диалектический материализм рассматривает математику, как и всю науку в целом, такой, как она есть, во всем богатстве и сложности ее связей и развития. И именно потому, что диалектический материализм стремится понять все богатство и всю сложность связей науки с действительностью, всю сложность ее развития, идущего от простого обобщения опыта к высшим абстракциям и от них к практике, именно потому, что самый свой подход к науке он постоянно приводит в соответствие с ее объективным содержанием, с ее новыми открытиями, именно поэтому и, в конечном счете только поэтому, он и оказывается единственной подлинно научной философией, ведущей к верному пониманию науки вообще и, в частности,— математики.

ЛИТЕРАТУРА

Статьи, посвященные общим вопросам математики

Колмогоров А. Н. Математика. БСЭ, т. 26.
Александров А. Д. Геометрия. БСЭ, т. 10.

Книги исторического характера

Башмакова И. Г. и Юшкевич А. П. Происхождение систем счисления. Энцикл. элемент. математики, т. I, Гостехиздат, 1951.

Гнеденко Б. В. Очерки по истории математики в России. Гостехиздат, 1946.

«Математика в СССР за тридцать лет (1917—1947)». Сборник обзорных статей, освещающих развитие отдельных отраслей математики в нашей стране. Гостехиздат, 1948.

Цейтен Г. Г. История математики в древности и в средние века. ГОНТИ, 1938.

Цейтен Г. Г. История математики в XVI и XVII веках. ГОНТИ, 1938.

Глава II

АНАЛИЗ

§ 1. ВВЕДЕНИЕ

Возникновение в конце средних веков новых производственных отношений в Европе, зарождение капитализма, шедшего тогда на смену феодальному строю, сопровождалось великими географическими открытиями и исследованиями. В 1492 г., основываясь на идее шарообразности Земли, Колумб открыл Новый Свет. Открытие Колумба широко раздвинуло рамки известного тогда мира и произвело переворот в умах людей. В конце 1400-х и начале 1500-х годов творили великие художники-гуманисты Леонардо да Винчи, Рафаэль, Микельанджело, обновляя искусство. В 1543 г. было опубликовано сочинение Коперника «Об обращении небесных кругов», совершенно видоизменившее лицо астрономии; в 1609 г. появилась «Новая астрономия» Кеплера, содержащая 1 и 2-й законы движения планет вокруг Солнца, а в 1618 г.— его книга «Гармонии мира», содержащая 3-й закон. Галилей, отправляясь от изучения творений Архимеда и смелых опытов, начал создавать новую механику, совершенно необходимую нарождавшейся тогда технике. В 1609 г. Галилей направил построенный им небольшой и несовершенный телескоп на ночное небо. Одного взгляда в телескоп было достаточно, чтобы разрушить идеальные небесные сферы Аристотеля и догмат о совершенстве небесных тел. Поверхность Луны оказалась покрытой горами и изрытой кратерами. Венера обнаружила фазы, как у Луны, Юпитер оказался окруженным четырьмя спутниками и давал как бы небольшую наглядную модель солнечяой системы. Млечный путь распался на отдельные звезды, и впервые была почувствована поражающе огромная удаленность звезд. Никогда научное открытие не производило такого впечатления на культурный мир[1].

Развитие дальнего мореплавания и вызванные им исследования астрономов, а также развитие новой техники и связанное с этим развитие механики сделали совершенно необходимым поиски способов решения мно-

[1] В основу этого абзаца положен текст прекрасной статьи академика С. И. Вавилова «Галилей» (БСЭ, т. 10, 1952).

жества возникавших тогда новых математических задач. Новизна этих задач состояла главным образом в том, что пришлось подвергать математическому изучению законы движения в широком смысле этого слова.

Природе чуждо состояние покоя и неподвижности. Вся природа, как отмечал Ф. Энгельс, начиная от мельчайших частиц ее до величайших тел, находится в вечном возникновении и уничтожении, в непрерывном течении, в неустанном движении и изменении. Каждая естественная наука изучает в конечном итоге те или иные стороны, те или иные формы этого движения. Математический анализ — раздел математики, дающий методы для *количественного* исследования разных процессов изменения, движения, зависимости одних величин от других. Неслучайно поэтому, что он возник в период, когда развитие механики и астрономии, побуждаемое вопросами техники и мореплавания, привело уже к достаточному накоплению наблюдений, измерений, гипотез и вплотную подвело науку к количественному исследованию простейших форм движения.

Само название «анализ бесконечно малых» ничего не говорит о предмете изучения, но зато подчеркивает метод, которым оперируют в этом отделе математики. Речь идет о специальном математическом методе бесконечно малых или в его современном виде — о методе пределов. Мы сразу приведем типичные примеры рассуждений, использующих метод пределов, а в одном из следующих параграфов уточним необходимые понятия.

Пример 1. Как экспериментально установил Галилей, путь s, проходимый свободно падающим в пустоте телом за время t, выражается формулой

$$s = \frac{gt^2}{2} \tag{1}$$

(g — постоянная величина, равная 9,81 м/сек^2)[1]. Какова же скорость падающего тела в каждой из точек его пути?

Пусть тело проходит через точку A в момент времени t. Рассмотрим, что произойдет затем за небольшой промежуток времени длительности Δt, т. е. за время от t до $t + \Delta t$. Пройденный путь получит некоторое приращение Δs. Прежний путь $s_1 = \frac{gt^2}{2}$; увеличенный путь

$$s_2 = \frac{g(t+\Delta t)^2}{2} = \frac{gt^2}{2} + \frac{g}{2}(2t\Delta t + \Delta t^2).$$

Отсюда находим приращение

$$\Delta s = s_2 - s_1 = \frac{g}{2}(2t\Delta t + \Delta t^2).$$

[1] Теперь формулу (1) выводят из общих законов механики, но исторически именно формула (1) была установлена экспериментально и послужила частицей опыта, обобщенного затем в этих законах.

Оно представляет собой путь, пройденный за время от t до $t + \Delta t$. Для нахождения средней скорости на участке пути Δs разделим Δs на Δt:

$$v_{\text{ср}} = \frac{\Delta s}{\Delta t} = gt + \frac{g}{2}\Delta t.$$

Заставляя Δt приближаться к нулю, мы будем получать средние скорости, неограниченно приближающиеся к истинной скорости в точке A. С другой стороны, мы видим, что второе слагаемое в правой части последнего равенства становится с уменьшением Δt исчезающе малым, так что $v_{\text{ср}}$ стремится при этом к величине gt, что принято записывать так:

$$v = \lim_{\Delta t \to 0} v_{\text{ср}} = \lim_{\Delta t \to 0} \frac{\Delta s}{\Delta t} = $$
$$= \lim_{\Delta t \to 0}\left(gt + \frac{g}{2}\Delta t\right) = gt.$$

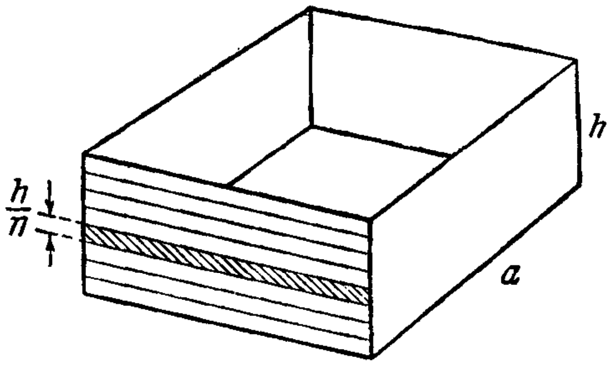


Рис. 1.

Следовательно, gt и есть истинная скорость в момент времени t.

Пример 2. Резервуар, имеющий квадратное дно со стороной a и вертикальные стенки высотой h, наполнен до краев водой (рис. 1). С какой общей силой давит вода на одну из стенок резервуара?

Разделим поверхность стенки на n горизонтальных полосок высотой $\frac{h}{n}$. Давление в каждой точке сосуда равно, как известно, напору вышележащего столба жидкости. Поэтому у нижнего края каждой из полосок давление, выраженное в соответствующих единицах, будет равно соответственно: $\frac{h}{n}, \frac{2h}{n}, \frac{3h}{n}, \ldots, \frac{(n-1)h}{n}, h$. Мы получим приближенное выражение искомой силы P, если будем считать давление в пределах каждой полоски постоянным. Таким образом, приближенное значение P равно

$$P \approx \frac{ah}{n}\cdot\frac{h}{n} + \frac{ah}{n}\cdot\frac{2h}{n} + \ldots + \frac{ah}{n}\frac{(n-1)h}{n} + \frac{ah}{n}h = \frac{ah^2}{n^2}(1+2+\ldots+n) =$$
$$= \frac{ah^2}{n^2}\cdot\frac{n(n+1)}{2} = \frac{ah^2}{2}\left(1+\frac{1}{n}\right).$$

Для отыскания истинной величины силы будем совершать разбиение на все более узкие полоски, неограниченно увеличивая n. С ростом n величина $\frac{1}{n}$ в последней формуле будет становиться все меньше и меньше, и в пределе мы получим точную формулу

$$P = \frac{ah^2}{2}.$$

Идея метода пределов проста и заключается в следующем. Чтобы определить некоторую величину, мы определяем сначала не ее самое, а некоторое ее приближение. При этом строится не одно приближение, а целый ряд приближений, все более и более точных. Затем из рассмотрения цепи этих приближений, т. е. из рассмотрения самого процесса приближения, определяется уже единственным образом точное значение величины. Этим по существу глубоко диалектическим методом устойчивое, постоянное познается как результат процесса, движения.

Математический метод пределов выработан в результате упорного труда многих поколений над задачами, которые не могли быть решены простыми приемами арифметики, алгебры и элементарной геометрии.

Каковы же были задачи, при решении которых складывались основные понятия анализа, какие приемы решения этих задач были созданы?

Математики 1600-х годов постепенно обнаружили, что большое число задач по исследованию разного рода движений и зависимостей одних величин от других, а также геометрических задач, не поддававшихся ранее решению, сводится к двум типам. Наиболее яркими и простыми примерами задач первого типа являются: вопрос нахождения скорости в данный момент при неравномерном движении и аналогичные ему вопросы о скорости изменения величин, а также задача о проведении касательной к кривым линиям. Эти задачи (к ним относится наш первый пример) привели к разделу анализа, получившему название «дифференциального исчисления». Простейшими примерами второго типа задач служат задачи на отыскание площади криволинейных фигур, суммарного пути, пройденного при неравномерном движении, вообще суммарного итога действия непрерывно изменяющейся величины (таким является второй из приведенных выше примеров). Эта группа задач привела к другому разделу анализа — интегральному исчислению. Так выделились две основные задачи: задача о касательных и задача о квадратурах.

В этой главе будет подробно описано, какие идеи были положены в основу решения обеих этих задач. Особенно важной оказалась теорема Ньютона и Лейбница о том, что задача о квадратурах есть в известном смысле обращение задачи о касательных. Для решения же задачи о касательных и для решения задач, которые к ней сводятся, был найден удобный и совершенно общий вычислительный алгорифм — общий способ, заведомо ведущий к решению,— метод производных, метод дифференцирования.

История создания и развития анализа и роль, которую в его зарождении сыграла созданная Декартом аналитическая геометрия, уже были описаны в главе I. Мы видим, что за вторую половину 1600-х и первую половину 1700-х годов произошло полное видоизменение всей математики. К имевшимся ее разделам — арифметике, элементарной геометрии, началам алгебры и тригонометрии — были добавлены такие общие методы, как аналитическая геометрия, дифференциальное исчисление и интегральное исчисление с теорией простейших дифференциальных уравнений. Теперь оказалось возможным решать задачи, о решении которых раньше нельзя было и мечтать.

Оказалось, что, если закон кривой линии не слишком сложный, то можно всегда построить к ней касательную в любой ее точке,— стоит лишь вычислить при помощи правил дифференциального исчисления так называемую производную, что во многих случаях потребует лишь нескольких минут. До того умели только проводить касательные к окружности и еще к двум-трем кривым и совершенно не ожидали, что есть общее решение задачи.

Если известен путь, проходимый точкой за любое данное время, то тем же способом можно сейчас же найти скорость точки в любой данный момент или ее ускорение. Обратно, по ускорению можно найти скорость и путь, применяя обращение дифференцирования — так называемое интегрирование. Отсюда нетрудно было, зная геометрические свойства эллипса, показать, что из ньютоновых законов движения и закона всемирного тяготения следует, что планеты должны двигаться вокруг Солнца по эллипсам согласно законам Кеплера.

Важнейшее значение в практике имеет вопрос о наибольших и наименьших величинах, так называемые задачи на максимум и минимум. Возьмем пример: из круглого бревна данного радиуса надо вытесать балку прямоугольного сечения такую, чтобы она была возможно более прочна на прогиб. С каким соотношением сторон ее следует сделать? Небольшое рассуждение о прочности балки прямоугольного сечения (с применением простых соображений интегрального исчисления), а затем решение задачи на максимум (для чего придется воспользоваться вычислением производной) дают ответ, что наибольшая прочность будет достигнута при прямоугольном сечении, высота которого относится к основанию, как $\sqrt{2}:1$. Задачи на максимум и минимум решаются так же просто, как задачи на проведение касательной.

В разных местах кривой линии, если она не прямая и не окружность, изогнутость кривой, вообще говоря, разная. Как вычислить радиус той окружности, которая так же изогнута, как данная линия около данной ее точки,— так называемый радиус кривизны кривой линии в данной ее точке? Оказывается, что и это столь же просто; надо только два раза применить операцию дифференцирования. Радиус кривизны играет большую роль во многих вопросах механики.

До изобретения новых исчислений умели находить только площадь многоугольников, круга, его сектора и сегмента и еще двух — трех фигур. Кроме того, еще Архимед дал способ вычислять площадь сегмента параболы. Способ, который он при этом употреблял, был основан на специальных свойствах параболы и был весьма остроумен. Это достижение Архимеда заставляло думать, что каждая новая задача на вычисление площади, пожалуй, потребует опять своих и еще более остроумных и трудных изысканий. Каково же было восхищение математиков, когда оказалось, что теорема Ньютона — Лейбница о том, что обращение задачи о касательных решает задачу о квадратурах, сейчас же делает возможным вычислять площади, ограниченные самыми различными кривыми. Выяснилось, таким образом, что есть общий способ, годный для бесчисленного множества самых разных фигур. То же касается вычисления объемов, поверхностей, длин кривых, масс неоднородных тел и т. д.

Еще большего достиг новый метод в механике. Казалось, не было такого вопроса механики, которого новые исчисления не осветили бы и не решили.

Незадолго перед тем Паскаль объяснил возрастание величины торричеллиевой пустоты с подъемом на гору, как следствие уменьшения атмосферного давления при подъеме. Но по какому закону уменьшается это давление? Вопрос сейчас же решается путем исследования простого дифференциального уравнения.

Морякам хорошо известно, что стоит обернуть канат вокруг причального шпиля один — два раза, и уже один человек может удержать большой корабль у причала. Почему это? Оказалось, что задача математически почти совершенно тождественна предыдущей и решается сразу.

Так за созданием анализа последовал период бурного развития его приложений в самых различных областях техники и естествознания. Созданный путем отвлечения от особенностей частных задач, математический анализ отражает тем самым совершенно реальные и глубокие свойства материального мира и именно потому становится средством исследования такого широкого круга вопросов. Механическое движение твердых тел, движение жидкостей и газов, движение их отдельных частиц и законы течения их масс, тепловые и электрические процессы, ход химических реакций и т. д., все эти явления исследуются соответствующими науками с широким использованием аппарата математического анализа.

Одновременно с расширением области своих приложений неизмеримо обогащается и сам анализ; возникают и совершенствуются такие его разделы, как теория рядов, приложения анализа к геометрии, теория дифференциальных уравнений.

У математиков середины 1700-х годов было очень распространено мнение, что любую задачу естествознания, если только удастся ее математически осмыслить, т. е. найти правильное ее математическое описание, можно решить при помощи аналитической геометрии, дифференциального и интегрального исчисления.

Постепенно начали приходить к более сложным задачам естествознания и техники, потребовавшим развития указанных методов. Для решения этих задач были последовательно созданы дальнейшие отделы математики: вариационное исчисление, теория функций комплексного переменного, теория поля, интегральные уравнения, функциональный анализ, которые решают эти новые классы задач. Но все эти новые исчисления по существу являются непосредственным продолжением и обобщением знаменитых исчислений, открытых в XVII веке. Величайшие математики XVIII века Д. Бернулли (1700—1782), Л. Эйлер (1707—1783) и Лагранж (1736—1813), прокладывая новые пути в науке, постоянно исходили из насущных задач точного естествознания. Энергичное развитие анализа продолжалось и в XIX веке, в котором жили и творили такие знаменитые математики, как Гаусс (1777—1855), Коши (1789—1857), М. В. Остроградский (1801—1861), П. Л. Чебышев (1821—1894), Риман (1826—1866), Абель (1802—1829), Вейерштрасс (1815—1897), внесшие весьма значительный вклад в развитие математического анализа.

Гениальный русский математик Н. И. Лобачевский также оказал влияние на развитие некоторых вопросов математического анализа.

Отметим еще крупнейших математиков, живших на рубеже XIX и XX века: А. А. Маркова (1856—1922), А. М. Ляпунова (1857—1918), А. Пуанкаре (1854—1912), Ф. Клейна (1849—1925), Д. Гильберта (1862—1943).

Во второй половине XIX века происходит глубокий критический пересмотр и уточнение самих основ анализа. Накопившиеся разнообразные и сильные методы анализа получают единое систематическое обоснование, соответствующее возросшему уровню математической строгости. Все это — методы, которыми, наряду с арифметикой, алгеброй, геометрией и тригонометрией, человек математически осмысливает окружающий его мир, описывает происходящие явления и решает связанные с ними важные для практики вопросы.

Особенно полного всестороннего развития и широких масштабов применения анализ достиг вместе со всей математической наукой у нас после Великой Октябрьской социалистической революции.

Аналитическая геометрия, дифференциальное, интегральное исчисления и теория дифференциальных уравнений изучаются во всех втузах, так что эти области математики известны в Советском Союзе миллионам людей; начальные сведения по этим наукам даются во многих техникумах; ставится вопрос об изучении их в средней школе.

В самое последнее время употребление больших быстродействующих математических вычислительных машин создает новый перелом в математике. Эти машины в соединении со всеми выше названными разделами математики дают новые необыкновенные возможности человеку.

Сейчас анализ с возникшими из него разделами представляет собой широко разветвленную математическую науку, состоящую из тесно связанных между собой обширных самостоятельных дисциплин. Каждая из них совершенствуется и движется вперед, причем значительная часть этих успехов принадлежит советским ученым.

Больше чем когда-либо раньше ведущей силой анализа являются запросы жизни, задачи, связанные с грандиозным развитием техники. На очереди стоят и постепенно с успехом решаются математические проблемы аэродинамики сверхзвуковых скоростей. Труднейшие задачи математической физики переходят сейчас в такую стадию, когда они смогут получать практические численные решения. В современной физике такие теории, как квантовая механика (а с ней связаны вопросы понимания явлений микромира), не только нуждаются для решения своих задач в самых высших частях современного математического анализа, но и не могли бы без его средств формулировать своих основных понятий.

Настоящая глава имеет целью в популярной форме познакомить читателя, знающего только элементарную математику, с возникновением и простейшими приложениями таких основных начальных понятий анализа, как функция, предел, производная и интеграл. Специальным разделам анализа будут посвящены другие главы этой книги. Соответственно глава эта носит более элементарный характер, и читатель, уже знакомый с обычным курсом анализа, может пропустить ее без ущерба для понимания дальнейшего.

§ 2. ФУНКЦИЯ

Понятие функции. В природе предметы и явления органически связаны между собой, зависят друг от друга. Устойчивые простейшие связи издавна изучались людьми. Знания о них накапливались и формулировались как физические законы. В массе случаев это были указания на то, что разные величины, количественно характеризующие некоторое явление, тесно связаны между собой, полностью обусловливаются одна значением другой. Например, размеры сторон прямоугольника вполне определяют его площадь, объем данного газа при определенной температуре обусловливается его давлением, удлинение данного металлического стержня определяется его температурой и т. п. Подобные закономерности и послужили источником понятия *функции*.

Уже в алгебраической формуле, позволяющей по каждому значению входящих в нее буквенных величин находить значение величины, выражаемой формулой, заложено понятие функции. Приведем примеры функций, заданных формулами.

1. Пусть в начальный момент времени материальная точка находилась в покое, а затем начала падать под воздействием силы тяжести. Тогда путь s, пройденный точкой за время t, выразится формулой

$$s = \frac{gt^2}{2}, \tag{1}$$

где g — ускорение силы тяжести.

2. Из квадрата со стороной a сделана открытая прямоугольная коробка высотой x (рис. 2). Объем V коробки будет вычисляться по формуле

$$V = x(a - 2x)^2. \tag{2}$$

Формула (2) позволяет для каждой высоты x, удовлетворяющей, очевидно, неравенству $0 \leqslant x \leqslant \frac{a}{2}$, найти объем коробки.

3. Пусть в центре круговой конькобежной дорожки врыт столб, и на

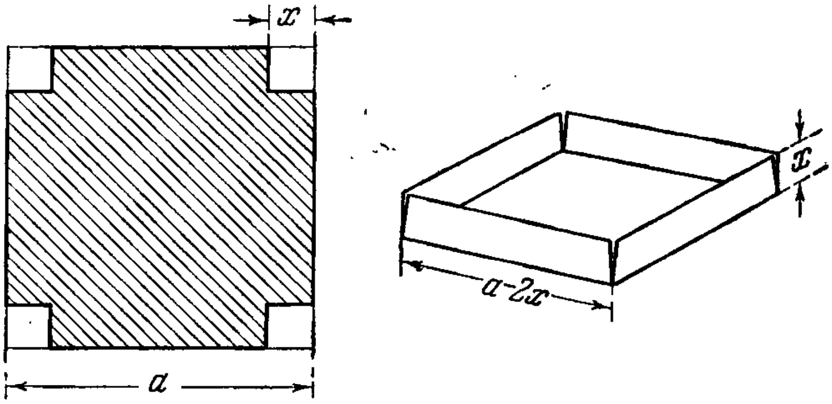


Рис. 2.

нем, на высоте h, подвешен фонарь (рис. 3). Освещенность T дорожки может быть выражена формулой

$$T = \frac{A \sin\alpha}{h^2 + r^2}, \tag{3}$$

где r — радиус дорожки, $\operatorname{tg}\alpha = \frac{h}{r}$, A — некоторая величина, характеризующая силу света фонаря. Зная высоту h, мы можем по формуле (3) вычислить T.

4. Корень квадратного уравнения

$$x^2 + px - 1 = 0 \tag{4}$$

вычисляется по формуле

$$x = -\frac{p}{2} \pm \sqrt{1 + \frac{p^2}{4}}. \tag{5}$$

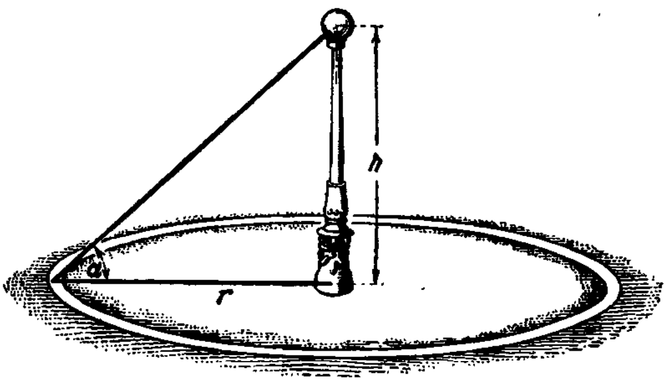


Рис. 3.

Характерным для формулы вообще и для приведенных выше формул в частности является то, что формула дает возможность по любому наперед заданному значению, которое может принимать одна величина (время t, высота x, высота h столба, коэффициент p уравнения), называемая независимой переменной, вычислять значения другой величины (путь s, объем V, освещенность T, корень x уравнения), которая носит название зависимой переменной или функции от первой величины.

Каждая из приведенных формул дает нам пример функции: путь s, пройденный точкой, есть функция времени t; объем V коробки есть функция ее высоты x; освещенность T дорожки есть функция высоты h столба; два корня квадратного уравнения (4) суть функции коэффициента p.

Следует отметить, что в одних случаях независимая переменная может принимать любые наперед заданные числовые значения, как это имеет место в примере 4, где коэффициент p квадратного уравнения (4), являющийся независимой переменной, может быть произвольным числом. В других случаях независимая переменная принимает любое значение из некоторого заранее определенного множества (совокупности) чисел, как в примере 2, где объем коробки есть функция от его высоты x, которая может принимать любое значение из множества чисел x, удовлетворяющих неравенствам $0 \leqslant x \leqslant \frac{a}{2}$. Точно так же в примере 3 освещенность T дорожки есть функция высоты столба s, которая может теоретически принимать любые значения, удовлетворяющие неравенству $s > 0$, а практически — любые значения s, удовлетворяющие неравенствам $0 < s \leqslant H$, где величина H определяется имеющимися в распоряжении администрации катка техническими возможностями.

Приведем еще такие примеры. Формула

$$y = \sqrt{1 - x^2}$$

определяет действительную функцию, выражающую соответствие между действительными числами x и y, очевидно, не для всех x, а только для тех из них, которые удовлетворяют неравенствам $-1 \leqslant x \leqslant 1$, а формула $y = \lg(1 - x^2)$ — для значений x, удовлетворяющих неравенствам $-1 < x < 1$.

Таким образом, приходится считаться с тем обстоятельством, что конкретные функции могут быть заданными не обязательно для всех возможных числовых значений, а только на некотором множестве числовых значений x, чаще всего заполняющих на оси x-ов некоторый отрезок (с концами или без концов).

Мы уже можем сейчас дать такое определение понятия функции, которое принято в математике в настоящее время.

Величина y есть функция от (независимой) величины x, если существует закон, в силу которого каждому значению x, принадлежащему к некоторому множеству чисел, соответствует определенное значение y.

Множество значений x, фигурирующее в этом определении, называется *областью определения функции.*

Каждое новое понятие часто порождает новую символику. Переход от арифметики к алгебре заключался в возможности построения формул, пригодных для любых числовых данных,— поиски общих решений привели к буквенной символике.

Задача анализа есть задача изучения функций — зависимостей одних величин от других; как в алгебре от конкретного числа переходят к произвольным числам — буквам, так и в анализе от конкретных формул мы переходим к произвольным функциям. Фразу «y есть функция от x» будем условно записывать так:

$$y = f(x).$$

Как в алгебре для разных чисел употребляются разные буквы, так и в анализе для обозначения различных зависимостей — функций — употребляются различные обозначения: $y = F(x)$, $y = \varphi(x)$, ...

Графики функций. Одной из наиболее плодотворных и блестящих идей второй половины XVII века является идея связи между понятием функции и геометрическим образом линии. Эта связь может быть осуществлена, например, посредством прямоугольной декартовой системы координат, с которой читатель в самых общих чертах, конечно, уже знаком из курса средней школы.

Зададим на плоскости прямоугольную декартову систему координат. Это значит, что мы выбираем на этой плоскости две взаимно перпендику-

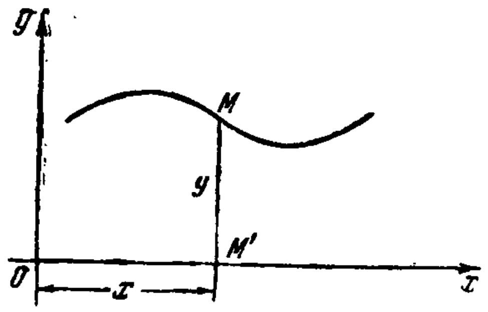


Рис. 4.

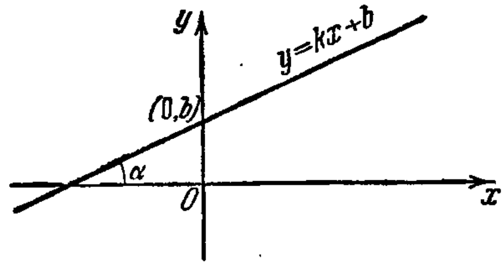


Рис. 5.

лярные прямые (ось абсцисс и ось ординат), на каждой из которых фиксировано положительное направление. Тогда каждой точке M плоскости можно поставить в соответствие два числа (x, y) — ее координаты, выражающие в выбранном масштабе соответственно расстояния точки M до оси ординат и до оси абсцисс, взятые с соответствующими знаками [1].

При помощи системы координат функции можно изобразить графически в виде некоторых линий. Пусть дана некоторая функция

$$y = f(x). \tag{6}$$

Это, как мы знаем, означает, что для каждого заданного x, принадлежащего к области определения данной функции, можно каким-либо способом определить, например, вычислить, соответствующее значение y. Будем придавать x всевозможные числовые значения. Для каждого x по нашему закону (6) определим y и построим в плоскости точку с координатами x и y. Таким образом, над каждой точкой M' оси x-ов (рис. 4) окажется расположенной точка M с координатами x и $y = f(x)$. Совокупность всех точек M образует некоторую линию, которую будем называть графиком нашей функции $y = f(x)$.

Итак, *графиком функции* $f(x)$ называется геометрическое место точек, координаты которых удовлетворяют уравнению (6).

В школе нас знакомили с графиками простейших функций. Так, читатель, вероятно, знает, что функция $y = kx + b$, где k и b — постоянные числа, графически изображается (рис. 5) прямой линией, образующей

[1] Число x называют абсциссой, а y — ординатой точки M.

с положительным направлением оси x-ов угол α, такой, что $\operatorname{tg}\alpha = k$, и пересекающей ось y-ов в точке $(0, b)$. Эта функция носит название *линейной функции*.

Линейные функции встречаются в приложениях особенно часто. Вспомним, что многие физические законы выражаются, и притом достаточно точно, линейными функциями. Например, длина l тела с хорошим приближением рассматривается как линейная функция его температуры

$$l = l_0 + \alpha l_0 t,$$

где α — коэффициент линейного расширения, l_0 — длина тела при $t = 0$. Если x есть время, а y — путь, пройденный за это время точкой, то линейная функция $y = kx + b$ выражает, очевидно, то обстоятельство, что точка движется равномерно со скоростью k; число же b обозначает расстояние нашей точки от места отсчета пути в момент времени $x_0 = 0$. Возможность приближенно считать равномерными различные изменения, хотя бы на малых участках, и простота линейной функции делают ее очень употребительной.

В других случаях необходимо применение иных функциональных зависимостей. Вспомним, например, закон Бойля-Мариотта

$$v = \frac{c}{p},$$

где зависимость между p и v состоит в обратной пропорциональности этих величин. График такой зависимости представляет собой гиперболу (рис. 6).

Сам физический закон Бойля-Мариотта соответствует случаю, когда p и v положительны; он описывается ветвью гиперболы, находящейся в первой четверти.

Случаи колебательных процессов сопровождаются периодическими движениями, которые в свою очередь описываются обычно тригонометрическими функциями, изменяющимися, как мы знаем, периодически. Например, если вывести из равновесия подвешенную пружину, растянув ее в пределах упругости, то ее точка A будет совершать вертикальные колебания, выражающиеся довольно точно законом

$$x = a\cos(pt + \alpha),$$

где x — отклонение точки A от положения равновесия, t — время, а числа a, p и α — некоторые постоянные, определяемые материалом, размерами и степенью начального растяжения пружины.

Надо иметь в виду, что функция может быть определена в различных областях различными формулами, и это может диктоваться обстоятельствами дела. Например, зависимость $Q = f(t)$ между температурой t одного грамма воды (льда) и количеством Q находящегося в нем тепла, когда t изменяется между $-10°$ и $+10°$, есть вполне определенная функция, которую затруднительно выразить единой формулой [1], но двумя форму-

[1] Это не значит, что такое представление невозможно. В главе XII (том 2) будет показано, как получить единую формулу.

лами эту функцию легко задать. Так как теплоемкость льда равна 0,5, а теплоемкость воды равна 1, то эта функция, если принять условно, что при $-10°$ величина $Q = 0$, выражается формулой

$$Q = 0{,}5t + 5,$$

когда t изменяется в промежутке $-10° \leqslant t < 0°$ и выражается другой формулой

$$Q = t + 85,$$

когда t изменяется в промежутке $0° < t \leqslant 10°$. При $t = 0$ эта функция оказывается неопределенной — многозначной; можно для удобства условиться,

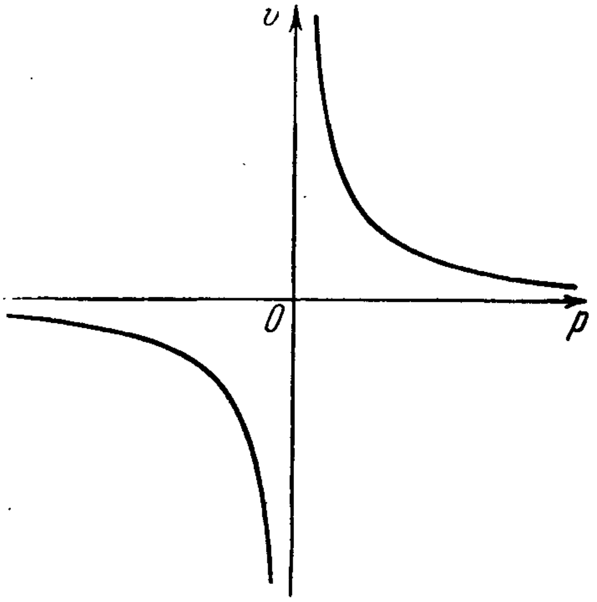


Рис. 6.

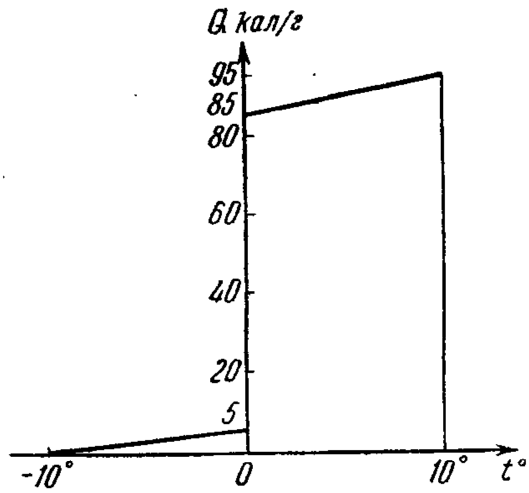


Рис. 7.

что при $t = 0$ она принимает одно вполне определенное значение, например $f(0) = 45$. График функции $Q = f(t)$ изображен на рис. 7.

Мы привели много примеров функций, заданных формулами. Способ задания функции при помощи формул с математической точки зрения является наиболее важным, так как при таком способе имеется наличие наиболее благоприятных условий для исследования свойств функции математическими методами.

Однако не нужно думать, что формула есть единственный способ задания функции. Существуют много других способов, среди них особое значение имеет график функции, дающий наглядное геометрическое ее изображение. Следующий пример может служить хорошей иллюстрацией этого.

Для того, чтобы узнать, как изменяется в течение суток температура воздуха, на метеорологических станциях пользуются прибором, называемым термографом. Термограф состоит из барабана, вращающегося вокруг своей оси при помощи часового механизма, и латунной изогнутой коробки, весьма чувствительной к изменениям температуры. При повышении температуры она разгибается, в результате этого прикрепленное к ней при помощи системы рычажков самопишущее перо поднимается вверх. Наоборот, понижение температуры влечет за собой опускание пера. На барабан навертывается соответствующим образом разграфленная бумажная лента, на которой перо вычерчивает непрерывную линию— график функции $T = f(t)$, выражающий зависимость между временем и температурой воздуха. При помощи полученного графика можно без вычислений определять значения температуры T для каждого момента времени t.

Приведенный пример показывает, что график сам по себе определяет функцию независимо от того, задана она формулой или нет.

Впрочем, мы вернемся к этому вопросу (см. главу XII, том 2) и покажем справедливость следующего весьма важного утверждения: каждый непрерывный график может быть представлен некоторой формулой или, как еще принято говорить, аналитическим выражением. Это верно и для многих разрывных графиков [1].

Отметим, что это утверждение, имеющее большое принципиальное значение, было полностью осознано в математике только в середине прошлого века. До того времени математики под термином «функция» понимали только аналитическое выражение (формулу). Но при этом они ошибочно думали, что далеко не всякому непрерывному графику соответствует аналитическое выражение, полагая, что раз уж функция задана формулой, то ее график должен обладать особенно хорошими свойствами сравнительно с другими графиками.

Однако в XIX веке было обнаружено, что все непрерывные графики могут быть заданы формулой, более или менее сложной. Этим исключительная роль аналитического выражения как способа определения функции была поколеблена, и в результате сформировалось новое, более гибкое определение понятия функции, которое было дано выше. По этому определению переменная y называется функцией от переменной x, если существует закон, в силу которого каждому значению x из области определения этой функции соответствует вполне определенное значение y, независимо от того, каким способом задан этот закон: формулой, графиком, таблицей или еще каким-либо другим способом.

Здесь уместно отметить, что в математической литературе высказанное определение часто связывают с именем математика Дирихле. Стоит подчеркнуть, что это определение было одновременно с Дирихле и независимо от него предложено Н. И. Лобачевским.

В заключение предлагаем в качестве упражнения нарисовать графики функций x^3, $\sqrt{x}$, $\sin x$, $\sin 2x$, $\sin\left(x+\frac{\pi}{4}\right)$, $\ln x$, $\ln(1+x)$, $|x-3|$, $\frac{x+|x|}{2}$.

[1] Конечно, высказанное утверждение будет вполне ясным читателю лишь после того, как будет точно определено, что, собственно, понимают в математике под терминами «формула» и «аналитическое выражение».

Следует отдать также себе отчет в том, что график функции, удовлетворяющей для всех значений x соотношению

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

симметричен относительно оси y-ов, а в случае соотношения

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

он симметричен относительно начала координат. Подумайте, как получить график функции $f(a+x)$, где a — постоянное число, из графика $f(x)$. Наконец, разберитесь в том, как, пользуясь графиками функций $f(x)$ и $\varphi(x)$, можно находить значения сложной функции $y = f[\varphi(x)]$.

§ 3. ПРЕДЕЛ

В § 1 было сказано, что современный математический анализ оперирует специальным методом, который вырабатывался на протяжении многих веков и который служит в анализе основным средством рассуждения. Речь здесь идет о методе бесконечно малых, или, что в сущности все равно, о методе пределов. Постараемся дать представление об этих понятиях. Для этого рассмотрим следующий пример.

Требуется вычислить площадь, ограниченную параболой, уравнение которой $y=x^2$, осью x-ов и прямой $x = 1$ (рис. 8). Элементарная математика не дает нам средств для решения этой задачи. Но вот как можно здесь поступить.

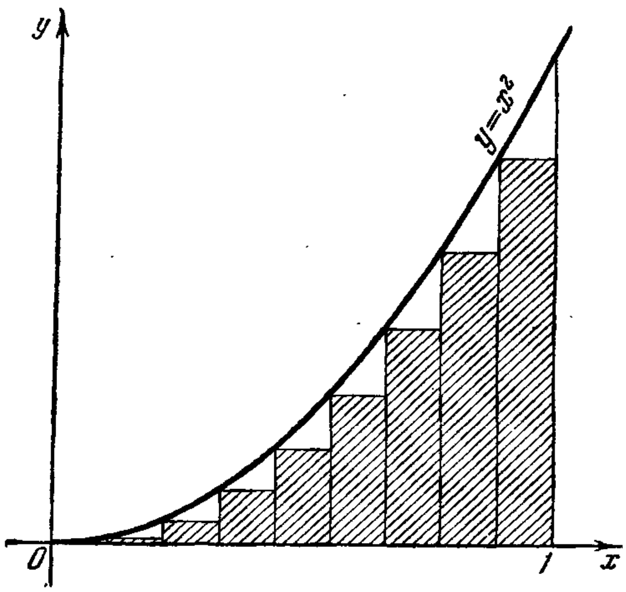


Рис. 8.

Разделим отрезок $[0, 1]$ оси x-ов на n равных частей точками

$$0, \frac{1}{n}, \frac{2}{n}, \ldots, \frac{n-1}{n}, 1$$

и построим на каждой из этих частей прямоугольник, левый верхний угол которого достигает параболы. В результате получим систему заштрихованных на рис. 8 прямоугольников, сумма S_n площадей которых равна

$$S_n = 0 \cdot \frac{1}{n} + \left(\frac{1}{n}\right)^2 \frac{1}{n} + \left(\frac{2}{n}\right)^2 \frac{1}{n} + \ldots + \left(\frac{n-1}{n}\right)^2 \frac{1}{n} = \frac{1^2 + 2^2 + \ldots + (n-1)^2}{n^3} =$$

$$= \frac{(n-1)\,n\,(2n-1)}{6n^3}\,[1].$$

[1] Если сложить отдельно левые и правые части очевидных равенств $(k+1)^3 - k^3 = 3k^2 + 3k + 1$, соответствующих различным $k = 1, 2, \ldots, n-1$, то получим уравнение $n^3 - 1 = 3\sigma_n + \frac{3(n-1)n}{2} + n - 1$ для $\sigma_n = 1^2 + 2^2 + \ldots + (n-1)^2$. Решив его, получим $\sigma_n = \frac{(n-1)\,n\,(2n-1)}{6}$.

Представим величину S_n в следующем виде:

$$S_n = \frac{1}{3} + \left(\frac{1}{6n^2} - \frac{1}{2n}\right) = \frac{1}{3} + \alpha_n. \tag{7}$$

Величина α_n, зависящая от n, хотя и имеет довольно громоздкий вид, но обладает замечательным свойством: если n неограниченно увеличивать, то α_n будет стремиться к нулю. Это свойство можно выразить еще так: если задать произвольное положительное число ε, то можно подобрать настолько большое значение N, что для всех n, бо́льших, чем N, число α_n будет по абсолютной величине меньше заданного ε[1].

Величина α_n есть пример бесконечно малой величины в том смысле, как этот термин понимается в современной математике.

Из рис. 8 мы видим, что если увеличивать число n неограниченно, то сумма S_n площадей заштрихованных прямоугольников будет стремиться к искомой площади криволинейной фигуры. С другой стороны, равенство (7) в силу того, что α_n стремится при неограниченном возрастании n к нулю, показывает, что сумма S_n при этом стремится к $^1/_3$. Отсюда следует, что искомая площадь S фигуры равна $^1/_3$, и мы нашу задачу решили.

Изложенный метод, таким образом, свелся к тому, что для нахождения некоторой величины S мы ввели другую, стремящуюся к ней *переменную* величину S_n, пробегающую отдельные частные значения S_1, S_2, S_3, ..., зависящие по некоторому закону от натуральных чисел $n = 1, 2, \ldots$ Затем после того, как было замечено, что переменная S_n может быть представлена в виде суммы постоянного числа $^1/_3$ и бесконечно малой α_n, мы заключили, что S_n стремится к $^1/_3$, и, таким образом, $S = {^1/_3}$. На языке современной теории пределов в данном случае можно сказать, что переменная величина S_n при неограниченном возрастании n стремится к пределу, равному $^1/_3$.

Дадим точное определение введенных понятий.

Если переменная величина α_n ($n = 1, 2, \ldots$) обладает тем свойством, что для всякого, сколь угодно малого положительного числа ε можно подобрать такое достаточно большое N, что для всех $n > N$ будет выпол-

---

[1] Например, если $\varepsilon = 0{,}001$, то можно считать $N = 500$. Действительно, так как

$$\frac{1}{6n^2} < \frac{1}{2n}$$

при целых положительных n, то

$$|\alpha_n| = \left|\frac{1}{6n^2} - \frac{1}{2n}\right| = \frac{1}{2n} - \frac{1}{6n^2} < \frac{1}{2n} < 0{,}001$$

для любого $n > 500$. Подобным образом можно было бы задать любые малые значения ε, например:

$$\varepsilon_1 = 0{,}0001, \ \varepsilon_2 = 0{,}00001, \ldots,$$

и к ним подобрать, как выше, соответствующие значения $N = N_1, N_2, \ldots$

пяться неравенство $|\alpha_n| < \varepsilon$, то говорят, что α_n есть *бесконечно малая величина*, и пишут

$$\lim_{n\to\infty} \alpha_n = 0 \quad \text{или} \quad \alpha_n \to 0,$$

С другой стороны, если какую-либо переменную x_n можно представить в виде суммы

$$x_n = a + \alpha_n,$$

где a — некоторое постоянное число, а α_n — бесконечно малая, то говорят, что переменная x_n при неограниченном возрастании n стремится к числу a, и пишут

$$\lim x_n = a \quad \text{или} \quad x_n \to a:$$

Число a называют *пределом* x_n. В частности, предел бесконечно малой величины есть, очевидно, нуль.

Рассмотрим примеры переменных величин

$$x_n = \frac{1}{n}, \ y_n = -\frac{1}{n^2}, \ z_n = \frac{(-1)^n}{n}, \ u_n = \frac{n-1}{n} = 1 - \frac{1}{n}; \ v_n = (-1)^n$$
$$(n = 1, 2, \ldots).$$

Очевидно, x_n, y_n и z_n являются бесконечно малыми, причем первая из них стремится к нулю убывая; вторая стремится к нулю возрастая, принимая все время отрицательные значения, а третья стремится к нулю, колеблясь около него. Далее $u_n \to 1$, а v_n не имеет вообще предела, так как с ростом n не приближается ни к какому постоянному числу, а все время колеблется, принимая значения то 1, то -1.

В анализе важное значение имеет также понятие *бесконечно большой* величины, которая определяется как переменная величина x_n $(n = 1, 2, \ldots)$, обладающая тем свойством, что для всякого сколь угодно большого положительного числа M можно указать такое N, что для всех $n > N$

$$|x_n| > M.$$

Тот факт, что величина x_n есть бесконечно большая, записывают так:

$$\lim x_n = \infty \quad \text{или} \quad x_n \to \infty.$$

Про такую величину x_n говорят, что она стремится к бесконечности. Если при этом она, начиная с некоторого значения, положительна (отрицательна), то это записывают так: $x_n \to +\infty$ $(x_n \to -\infty)$. Например, при $n = 1, 2, \ldots$

$$\lim n^2 = +\infty, \ \lim(-n^3) = -\infty;$$

$$\lim \lg\frac{1}{n} = -\infty, \ \lim \operatorname{tg}\left(\frac{\pi}{2} + \frac{1}{n}\right) = -\infty.$$

Легко видеть, что если величина α_n есть бесконечно большая, то $\beta_n = \frac{1}{\alpha_n}$ есть бесконечно малая, и наоборот.

Две переменные величины x_n и y_n можно складывать, вычитать, умножать и делить друг на друга и получать новые, вообще говоря, переменные величины: сумму $x_n + y_n$, разность $x_n - y_n$, произведение $x_n y_n$ и частное $\frac{x_n}{y_n}$. Их частными значениями будут соответственно

$$x_1 \pm y_1, \quad x_2 \pm y_2, \quad x_3 \pm y_3, \ldots$$

$$x_1 y_1, \quad x_2 y_2, \quad x_3 y_3, \ldots$$

$$\frac{x_1}{y_1}, \quad \frac{x_2}{y_2}, \quad \frac{x_3}{y_3}, \ldots$$

Можно также доказать сам по себе довольно очевидный факт, что если переменные x_n и y_n стремятся к конечным пределам, то их сумма, разность, произведение и частное также стремятся к пределам, равным соответственно сумме, разности, произведению и частному этих пределов. Это можно записать так:

$$\lim (x_n \pm y_n) = \lim x_n \pm \lim y_n; \quad \lim (x_n y_n) = \lim x_n \lim y_n;$$

$$\lim \frac{x_n}{y_n} = \frac{\lim x_n}{\lim y_n}.$$

Только в случае частного надо предполагать, что предел знаменателя ($\lim y_n$) не равен нулю. Если $\lim y_n = 0$, а $\lim x_n \neq 0$, то отношение x_n к y_n стремится уже не к конечному пределу, а к бесконечности.

Весьма интересный и в то же время важный случай будет иметь место, если и числитель и знаменатель стремятся одновременно к нулю. В этом случае нельзя заранее сказать, будет ли отношение $\frac{x_n}{y_n}$ стремиться к пределу, и если будет, то к какому, поскольку ответ на этот вопрос всецело зависит от характера стремления x_n и y_n к нулю. Например, если

$$x_n = \frac{1}{n}, \quad y_n = \frac{1}{n^2}, \quad z_n = \frac{(-1)^n}{n} \quad (n = 1, 2, \ldots),$$

то

$$\frac{y_n}{x_n} = \frac{1}{n} \to 0, \qquad \frac{x_n}{y_n} = n \to \infty.$$

С другой стороны, величина

$$\frac{x_n}{z_n} = (-1)^n,$$

очевидно, не стремится ни к какому пределу.

Таким образом, случай, когда числитель и знаменатель дроби оба стремятся к нулю, не может быть предусмотрен заранее общими теоремами, и для каждой отдельной дроби такого рода требуется проводить специальное исследование.

В дальнейшем мы увидим, что основная задача дифференциального исчисления, которую можно рассматривать как задачу нахождения скорости неравномерного движения в данный момент, сводится к определению предела отношения двух бесконечно малых величин — приращения пути к приращению времени.

Выше мы рассматривали переменные x_n, принимающие последовательность числовых значений x_1, x_2, x_3, ..., x_n, ..., в то время как значок n неограниченно пробегает ряд натуральных чисел $n = 1, 2, 3, \ldots$ Но можно было бы считать, что n изменяется непрерывно, как время, и при этом условии определить по аналогии предел переменной x_n. Свойства таких пределов совершенно аналогичны свойствам, сформулированным выше для дискретных (т. е. не непрерывных) переменных. Заметим еще, что здесь не имеет никакого значения то обстоятельство, что n неограниченно возрастает. С тем же эффектом можно считать, что n, непрерывно изменяясь, приближается к некоторому заданному значению n_0.

В качестве примера проследим за изменением величины $\frac{\sin x}{x}$, когда x приближается к нулю. Вот значения этой величины при нескольких значениях x:

x	$\frac{\sin x}{x}$
0,50	0,9589 . . .
0,10	0,9983 . . .
0,05	0,9996 . . .
. . .	. . .

(значения x предполагаются выраженными в радианной мере).

Повидимому, с приближением x к нулю величина $\frac{\sin x}{x}$ стремится к 1, но это, конечно, еще надо строго доказать. Доказательство можно получить, например, исходя из следующего неравенства, справедливого для всех отличных от нуля углов первой четверти:

$$\sin x < x < \operatorname{tg} x.$$

Разделив все части неравенства на $\sin x$, получим

$$1 < \frac{x}{\sin x} < \frac{1}{\cos x},$$

откуда следует

$$\cos x < \frac{\sin x}{x} < 1.$$

Но так как с уменьшением x до нуля $\cos x$ стремится к 1, то и величина $\frac{\sin x}{x}$, заключенная между $\cos x$ и 1, стремится к 1, т. е.

$$\lim_{x\to 0}\frac{\sin x}{x}=1.$$

Мы будем иметь случай использовать это обстоятельство.

Наше равенство доказано для случая, когда x стремится к нулю, оставаясь положительным. Видоизменяя очевидным образом доказательство, можно его получить и для случая, когда x стремится к нулю, принимая отрицательные значения.

Теперь остановимся еще на одном вопросе. Переменная величина может иметь и не иметь предела. Возникает вопрос, нельзя ли дать критерий, при помощи которого можно установить существование предела у переменной. Остановимся на одном важном и достаточно общем случае, когда такой критерий может быть дан. Представим себе, что переменная величина x_n возрастает или по крайней мере не убывает, т. е. выполняются неравенства

$$x_1\leqslant x_2\leqslant x_3\leqslant\ldots,$$

и пусть, кроме того, мы обнаружили, что все значения ее не превышают одного и того же числа M, т. е. $x_n\leqslant M$ $(n=1, 2,\ldots)$. Если значения x_n и число M отметить на оси x-ов, то получится, что переменная точка x_n движется по оси вправо, оставаясь, однако, все время левее точки M. Довольно очевидно, что переменная точка x_n должна обязательно стремиться к некоторой предельной точке a, которая находится либо левее M, либо, в крайнем случае, совпадает с M.

Таким образом, в рассматриваемом случае существует предел

$$\lim x_n=a$$

нашей переменной.

Приведенное рассуждение носит наглядный характер, однако не может рассматриваться как доказательство. В современных курсах математического анализа дается полное обоснование этого факта на основе теории действительного числа.

В качестве примера рассмотрим переменную

$$u_n=\left(1+\frac{1}{n}\right)^n \qquad (n=1, 2, 3, \ldots).$$

Первые ее значения $u_1=2$, $u_2=2{,}25$, $u_3\approx 2{,}37$, $u_4\approx 2{,}44$, ..., как мы видим, возрастают. Можно, разлагая наше выражение по биному Ньютона, показать, что возрастание имеет место при любом значении n. Больше того, можно легко показать, что для всех n справедливо неравенство $u_n<3$. В таком случае наша переменная обязательно имеет предел, не превышающий числа 3. Как мы увидим дальше, этот предел играет очень большую роль в математическом анализе, являясь в известном смысле наиболее естественным основанием для логарифмов чисел.

Этот предел принято обозначать буквой e. Он равен

$$e = \lim_{n \to \infty}\left(1 + \frac{1}{n}\right)^n = 2{,}718281828459045\ldots$$

Более детальные исследования показывают, что число e не является рациональным [1].

Можно также показать, что рассматриваемый предел существует и равен e не только когда $n \to +\infty$, но и когда $n \to -\infty$. При этом n в обоих случаях может пробегать не только целые числа.

Мы еще остановимся на одной важной роли понятия предела в естествознании. Она заключается в том замечательном факте, что только при помощи понятия предела (предельного перехода) нам удается давать полное исчерпывающее определение многих встречающихся в естествознании конкретных величин.

Рассмотрим пока следующий геометрический пример. В курсе школьной геометрии сначала изучаются фигуры, ограниченные прямолинейными отрезками. Затем встает более трудная задача о нахождении длины окружности данного радиуса.

Если проанализировать трудности, стоящие в связи с решением этой задачи, то они сводятся к следующему.

Надо отдать себе отчет в том, чтó такое длина окружности, т. е. дать точное ее определение. Существенно, чтобы определение сводилось к длинам прямолинейных отрезков, а также чтобы оно давало возможность эффективно вычислить длину окружности.

Само собой разумеется, что результат вычисления должен согласовываться с практикой. Если бы мы, например, взяли окружность, сделанную из реальной нити, то, разрезав ее и вытянув, мы должны были бы получить отрезок, длина которого в пределах точности измерений должна совпадать с вычисленной.

Как известно из школьного курса, решение этой задачи сводится к следующему определению. Длиной окружности называется предел, к которому стремится периметр вписанного в нее правильного[2] многоугольника при неограниченном увеличении числа его сторон. Таким образом, решение поставленной задачи существенно базируется на понятии предела.

Подобным образом определяется длина произвольной гладкой кривой. В ближайших параграфах мы узнаем ряд примеров геометрических

---

[1] В связи с этим следует отметить, что сложение, вычитание, умножение и деление (если исключить деление на нуль) рациональных чисел, т. е. чисел вида $\frac{p}{q}$ где, p и q — целые, приводит снова к числам рациональным. Но это не так в отношении операции предела. Предел последовательности рациональных чисел может быть числом иррациональным.

[2] Правильность многоугольника не имеет значения. Здесь важно только, чтобы наибольшая сторона переменного вписанного в окружность многоугольника стремилась к нулю.

и физических величин, точное определение которых может быть дано лишь благодаря применению понятия предела.

Понятия предела и бесконечно малой окончательно сформировались в начале прошлого века. Их определения, приведенные выше, связаны с именем Коши, до которого в математике оперировали с менее четкими понятиями. Современное представление о пределе, о бесконечно малой как переменной величине, о действительных числах явилось результатом развития математического анализа, средством оформления и закрепления его успехов.

§ 4. НЕПРЕРЫВНЫЕ ФУНКЦИИ

Непрерывные функции образуют основной класс функций, с которыми оперирует математический анализ. Представление о непрерывной функции можно получить, если сказать, что график ее непрерывен, т. е. его можно начертить, не отрывая карандаша от бумаги.

Непрерывная функция математически выражает одно свойство, с которым нам приходится часто встречаться на практике, заключающееся в том, что малому приращению независимой переменной соответствует малое же приращение зависимой от нее переменной (функции). Прекрасными примерами непрерывной функции могут служить различные законы движения тел $s = f(t)$, выражающие зависимости пути s, пройденного телом, от времени t. Время и пространство непрерывны, при этом тот или иной закон движения тела $s = f(t)$ устанавливает между ними определенную непрерывную связь, характеризующуюся тем, что малому приращению времени соответствует малое же приращение пути.

К абстракции непрерывности человек пришел, наблюдая окружающие его, так называемые сплошные среды — твердые, жидкие или газообразные, например металлы, воду, воздух. На самом деле, как теперь хорошо известно, всякая физическая среда представляет собой скопление большого числа отделенных друг от друга движущихся частиц. Однако эти частицы и расстояния между ними настолько малы по сравнению с объемами сред, с которыми приходится иметь дело в макроскопических физических явлениях, что многие такие явления можно достаточно хорошо изучать, если считать приближенно массу изучаемой среды без всяких просветов, непрерывно распределенной в занятом ею пространстве. На таком допущении базируются многие физические дисциплины, например гидродинамика, аэродинамика, теория упругости. Математическое понятие непрерывности играет, естественно, в этих дисциплинах, как и во многих других, большую роль.

Рассмотрим какую-либо функцию $y = f(x)$ и вполне определенное значение независимой переменной x_0. Если наша функция отражает некоторый непрерывный процесс, то значениям x, мало отличающимся от x_0, должны соответствовать значения функции $f(x)$, мало отличающиеся от значения $f(x_0)$ в точке x_0. Таким образом, если приращение $x - x_0$

независимой переменной мало, то должно быть малым также и соответствующее приращение $f(x) - f(x_0)$ функции. Иными словами, если приращение независимой переменной $x - x_0$ стремится к нулю, то приращение $f(x) - f(x_0)$ функции должно, в свою очередь, стремиться к нулю, что может быть записано следующим образом:

$$\lim_{x-x_0 \to 0} [f(x) - f(x_0)] = 0. \tag{8}$$

Это соотношение и является математическим определением непрерывности функции в точке x_0.

Функция $f(x)$ называется *непрерывной в точке* x_0, если выполняется равенство (8).

Дадим еще такое определение:

Функция называется *непрерывной для всех значений, принадлежащих к данному отрезку*, если она непрерывна в каждой точке x_0 этого отрезка, т. е. в каждой такой точке выполняется равенство (8).

Таким образом, для того чтобы ввести математическое определение свойства функции, заключающегося в том, что график ее есть непрерывная (в обычном понимании этого термина) кривая, появилась необходимость определить сначала локальное, местное свойство непрерывности (непрерывность в точке x_0), а затем на этой основе определить непрерывность функции на целом отрезке.

Приведенное определение, впервые указанное в начале прошлого столетия Коши, является общепринятым в современном математическом анализе. Проверка на многочисленных конкретных примерах показала, что это определение хорошо соответствует сложившемуся у нас практическому представлению о непрерывной функции, например представлению о непрерывном графике.

В качестве примеров непрерывных функций могут служить известные читателю из школьной математики элементарные функции x^n, $\sin x$, $\cos x$, a^x, $\lg x$, $\arcsin x$, $\arccos x$. Все перечисленные функции непрерывны на отрезках изменения x, где они определены.

Если непрерывные функции складывать, вычитать, умножать и делить (при знаменателе, не равном нулю), то в результате мы снова придем к непрерывной функции. Однако при делении непрерывность, как правило, нарушается для тех значений x_0, при которых функция, стоящая в знаменателе, обращается в нуль. Результат деления представляет собой тогда *разрывную* в точке x_0 функцию.

Функция $y = \frac{1}{x}$ может служить примером разрывной в точке $x = 0$ функции. Ряд других примеров разрывных функций дают графики, изображенные на рис. 9.

Мы рекомендуем читателю внимательно рассмотреть эти графики. Отметим, что разрывы функций бывают разные: иногда с приближением x к точке x_0, где функция претерпевает разрыв, предел $f(x)$ существует,

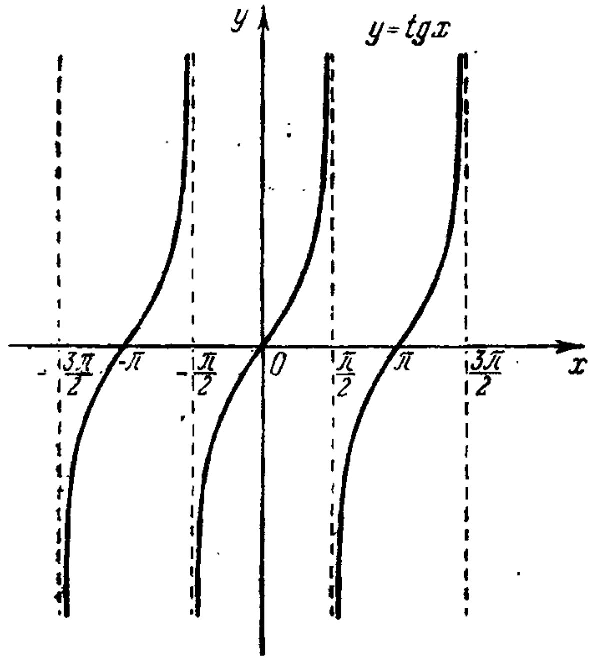


Рис. 9 а.

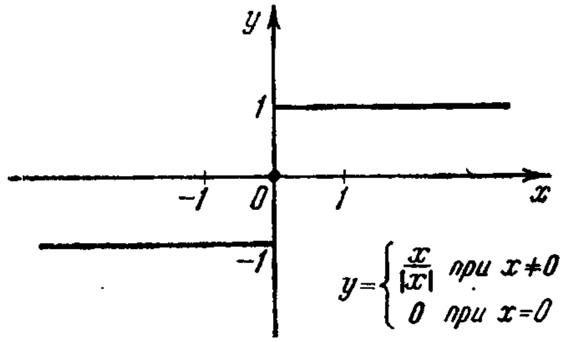


Рис. 9 б.

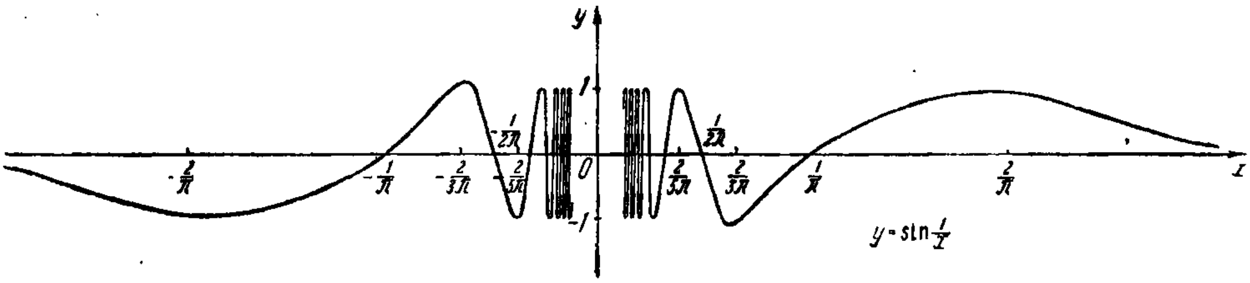


Рис. 9 в.

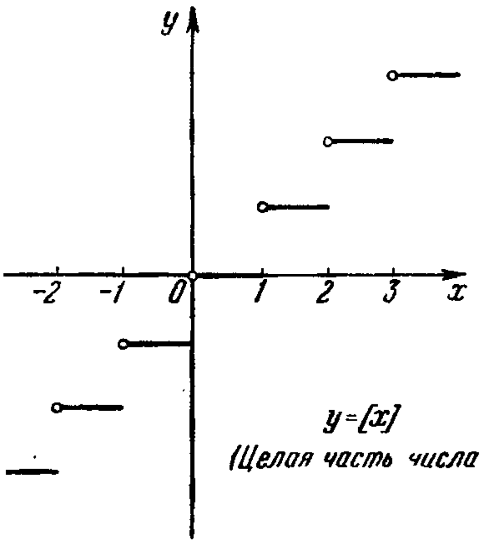


Рис. 9 г.

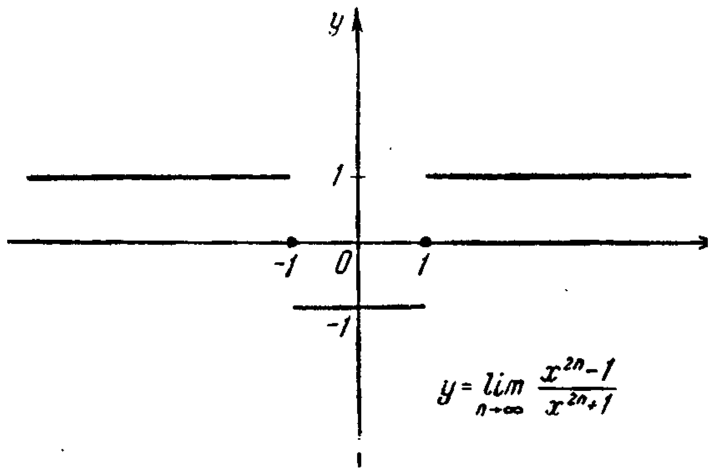


Рис. 9 д.

но отличен от $f(x_0)$, а иногда, как на рис. 9 в, этого предела просто не существует. Бывает и так, что с приближением x к x_0 с одной стороны $f(x)-f(x_0)\to 0$, а если $x\to x_0$, приближаясь с другой стороны, то $f(x)-f(x_0)$ уже не стремится к нулю. В этом случае, конечно, мы имеем разрыв функции, хотя про нее можно сказать, что она в этой точке «непрерывна с одной стороны». Все эти случаи можно проследить на приведенных графиках.

Для упражнения предложим читателю самому ответить на вопрос, чему должны равняться функции $\frac{\sin x}{x}$, $\frac{1-\cos x}{x^2}$, $\frac{x^3-1}{x-1}$, $\frac{\operatorname{tg} x}{x}$ в точках, где они не заданы (в точках, где знаменатель обращается в нуль), чтобы они были непрерывны в этих точках, и можно ли указать такие числа для функций $\operatorname{tg} x$, $\frac{1}{x-1}$, $\frac{x-2}{(x^2-4)}$?

Разрывные функции в математике отражают многочисленные скачкообразные процессы, встречающиеся в природе. При ударе, например, величина скорости тела меняется скачкообразно. Многие качественные переходы сопровождаются скачками. В § 2 мы приводили пример функции $Q=f(t)$, выражавшей зависимость количества тепла, находящегося в данном количестве воды (или льда), от температуры. Вблизи температуры таяния льда количество тепла $Q=f(t)$ с изменением t меняется скачкообразно.

Функции с отдельными разрывами, наряду с непрерывными функциями, довольно часто встречаются в анализе. Примером функции более сложной, с бесконечным числом разрывов, может служить так называемая функция Римана, равная нулю во всех иррациональных точках и равная $\frac{1}{q}$ в рациональных точках вида $x=\frac{p}{q}$ (где $\frac{p}{q}$ — несократимая дробь). Эта функция разрывна во всех рациональных точках и непрерывна в иррациональных. Несколько изменив ее, можно без труда получить пример функции, разрывной уже во всех точках[1]. Заметим кстати, что даже для таких сложных функций современный анализ обнаруживает много интересных закономерностей, которые исследуются в одной из самостоятельных ветвей анализа — теории функций действительного переменного. Большой вклад в эту чрезвычайно быстро развивавшуюся за последние 50 лет теорию внесен советскими математиками, особенно Московской школой теории функций.

§ 5. ПРОИЗВОДНАЯ

Следующим основным понятием анализа является понятие *производной*. Разберем две задачи, из решения которых оно исторически возникло.

Скорость. Во введении к этой главе мы уже определяли скорость свободно падающего тела. При этом мы пользовались предельным переходом

[1] Достаточно положить ее равной не нулю, а, скажем, единице в иррациональных точках.

от средней скорости на малых участках пути к скорости в данном месте в данный момент. Тот же прием может быть положен в основу определения мгновенной скорости при любом неравномерном движении. В самом деле, пусть функция

$$s = f(t) \tag{9}$$

выражает зависимость пройденного материальной точкой пути s от времени t. Чтобы найти скорость в момент $t = t_0$, рассмотрим некоторый промежуток времени от t_0 до $t_0 + h$ $(h \neq 0)$. За это время точка пройдет путь

$$\Delta s = f(t_0 + h) - f(t_0).$$

Средняя скорость $v_{\text{ср}}$ на этом участке будет зависеть от h

$$v_{\text{ср}} = \frac{\Delta s}{h} = \frac{1}{h}\{f(t_0 + h) - f(t_0)\},$$

и тем точнее будет выражать истинную скорость в момент t_0, чем меньше будет h. Отсюда следует, что истинная скорость в момент t_0 равна пределу

$$v = \lim_{h \to 0} \frac{f(t_0 + h) - f(t_0)}{h}$$

отношения приращения пути к приращению времени, когда последнее стремится к нулю, оставаясь все время не равным ему.

Чтобы вычислять скорость при разных законах движения, остается научиться находить этот предел для разных функций $f(t)$.

Касательная. К отысканию совершенно аналогичного предела приводит другая, на этот раз геометрическая, задача — о проведении касательной к произвольной плоской кривой.

Пусть кривая C есть график функции $y = f(x)$ и A — точка на кривой C с абсциссой x_0 (рис. 10). Какую прямую называть касательной к C в точке A? В элементарной геометрии этот вопрос не ставится. Для единственной изученной там кривой линии — окружности — касательная определялась как прямая, имеющая с окружностью только одну общую точку. Но для других кривых подобное определение явно не будет соответствовать наглядному представлению о «касании». Так, из двух прямых L и M, изображенных на рис. 11, первая, очевидно, не касается начерченной кривой (синусоиды), хотя имеет с ней всего одну общую точку, а вторая имеет с кривой много общих точек, но тем не менее касается кривой в каждой из них.

Чтобы дать определение касательной, рассмотрим на кривой C (рис. 10) еще одну, отличную от A, точку A' с абсциссой $x_0 + h$. Проведем секущую AA' и обозначим угол, который она образует с осью x-ов, через β. Будем теперь приближать точку A' по кривой C к A. Если при этом секущая AA' будет стремиться к некоторому предельному положению, то прямую T, имеющую это предельное положение, и называют *касательной* в точке A. Очевидно, угол α, который образует с осью x-ов прямая T, должен равняться пределу переменного угла β.

Величину $\operatorname{tg}\beta$ легко определить из треугольника ABA' (рис. 10):

$$\operatorname{tg}\beta=\frac{BA'}{AB}=\frac{f(x_0+h)-f(x_0)}{h}.$$

Для предельного положения должно быть

$$\operatorname{tg}\alpha=\lim_{A'\to A}\operatorname{tg}\beta=\lim_{h\to 0}\frac{f(x_0+h)-f(x_0)}{h},$$

т. е. тангенс угла наклона касательной равен пределу отношения приращения функции $f(x)$ в точке x_0 к соответствующему приращению независимой переменной, когда последнее стремится к нулю, оставаясь все время не равным ему.

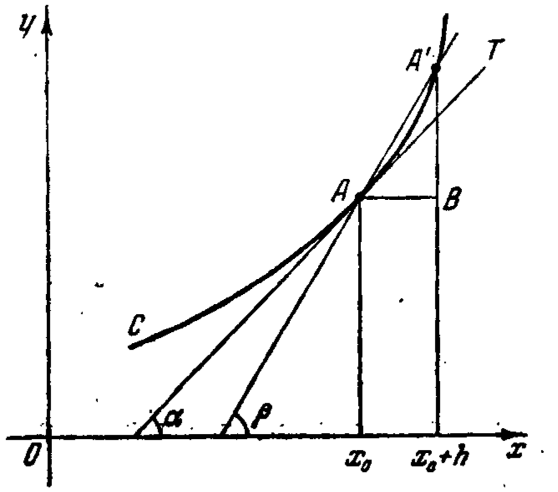


Рис. 10.

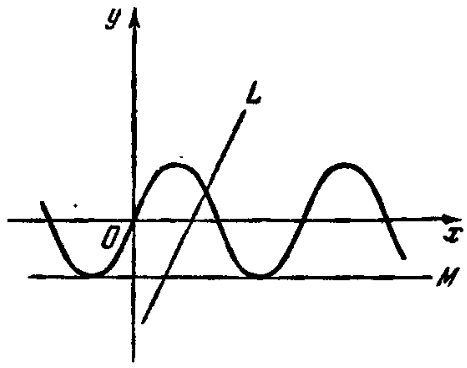


Рис. 11.

Вот еще пример, сводящийся к отысканию аналогичного предела. По проводу течет ток переменной силы. Допустим, что известна функция $Q=f(t)$, выражающая количество электричества, прошедшее через фиксированное сечение провода за время t. За период от t_0 до t_0+h через это сечение протекает количество электричества ΔQ, равное $f(t_0+h)-f(t_0)$. Средняя сила тока при этом равна

$$I_{\text{ср}}=\frac{\Delta Q}{h}=\frac{f(t_0+h)-f(t_0)}{h}.$$

Предел этого отношения при $h\to 0$ даст нам силу тока в момент t_0

$$I=\lim_{h\to 0}\frac{f(t_0+h)-f(t_0)}{h}.$$

Все три рассмотренные нами задачи, несмотря на то, что они относятся к различным областям человеческого знания: механике, геометрии, теории электричества, привели к одной и той же математической операции, которую нужно произвести над некоторой функцией. Надо найти предел отношения приращения функции к соответствующему приращению h независимой переменной при $h\to 0$. Мы могли бы как угодно увеличить число самых разнообразных задач, решение которых сводится к подобной операции. К ней, например, приводит вопрос о скорости химической реакции, о плотности неравномерно распределенной массы и др. Ввиду исключительной роли, которую играет эта операция над функцией, она получила особое название — операция дифференцирования функции. Результат этой операции носит название *производной.*

Итак, *производная от функции* $y = f(x)$, точнее, *значение производной в данной точке* x, есть предел [1], к которому стремится отношение приращения $f(x+h) - f(x)$ функции к приращению h независимой переменной, когда последнее стремится к нулю. Часто обозначают $h = \Delta x$, $f(x + \Delta x) - f(x) = \Delta y$, тогда определение производной записывается кратко: $\lim\limits_{\Delta x \to 0} \frac{\Delta y}{\Delta x}$.

Значение производной очевидно зависит от точки x, в которой оно найдено. Поэтому производная функции $y = f(x)$ есть в свою очередь некоторая функция от x. Производную принято обозначать так:

$$f'(x) = \lim_{h \to 0} \frac{f(x+h) - f(x)}{h} = \lim_{\Delta x \to 0} \frac{\Delta y}{\Delta x}.$$

Отметим еще другие принятые обозначения производной:

$$\frac{df(x)}{dx}, \text{ или } \frac{dy}{dx}, \text{ или } y', \text{ или } y'_x.$$

Следует заметить, что обозначение $\frac{dy}{dx}$ в самой записи выглядит как дробь, хотя читается как единый знак производной. В следующих параграфах числитель и знаменатель этой «дроби» приобретут для нас самостоятельный смысл, причем их отношение как раз будет совпадать с производной, что вполне оправдывает такую запись.

Результаты рассмотренных выше примеров теперь можно сформулировать так.

Скорость точки, для которой пройденный путь s есть данная функция времени $s = f(t)$, равна производной этой функции:

$$v = s' = f'(t).$$

Короче: скорость есть производная пути по времени.

Тангенс угла наклона касательной к кривой $y = f(x)$ в точке с абсциссой x равен производной функции $f(x)$ в этой точке:

$$\operatorname{tg} \alpha = y' = f'(x).$$

Сила тока I в момент t, если $Q = f(t)$ выражает количество электричества, прошедшее за время t через сечение провода, равна производной

$$I = Q' = f'(t).$$

Сделаем следующее замечание. Скорость неравномерного движения в данный момент — это чисто физическое понятие, возникшее из практики. Человек пришел к нему в результате многочисленных наблюдений над различными конкретными движениями. Изучение неравномерного дви-

[1] Разумеется, если указанный предел существует. В противном случае говорят, что функция в рассматриваемой точке x не имеет производной.

жения тела на различных участках его пути, сравнение различных таких движений, происходящих одновременно, в частности, изучение явлений столкновений тел, — все это составляет практический материал, приведший к созданию физического понятия скорости неравномерного движения в данный момент. Но точное определение скорости необходимо должно включать в себя способ определения ее численной величины. Это оказывается возможным сделать при помощи понятия производной.

В механике, по определению, величина скорости тела, движущегося по закону $s = f(t)$ в момент времени t, полагается равной производной от функции $f(t)$ для значения t.

Рассуждения в начале этого параграфа, с одной стороны, показали целесообразность введения операции получения производной, а с другой — разумно обосновали сформулированное выше определение скорости в данный момент.

Таким образом, когда мы ставили вопрос об отыскании скорости движущейся неравномерно точки, мы, собственно говоря, имели только опытное представление о ее величине, но точного определения не имели. В результате соответствующего анализа мы пришли к точному определению величины скорости в данный момент — это производная от пути по времени. Полученный результат имеет весьма важное практическое значение, так как на основе этого определения наше опытное представление о скорости обогатилось еще возможностью ее вычисления.

Сказанное, разумеется, относится и к силе тока и ко многим другим понятиям, выражающим скорость того или иного процесса (физического, химического и т. п.).

Обстоятельство, которое мы сейчас отметили, может служить примером многочисленных фактов подобного рода, когда практика приводит к определенному понятию, имеющему реальный смысл (скорость, работа, плотность, площадь и т. д.), а математика помогает это понятие четко определить, после чего мы получаем возможность оперировать данным понятием в нужных нам расчетах.

Мы уже отмечали в начале главы, что понятие производной возникло прежде всего как результат многовековых усилий, направленных к решению задачи о проведении касательной к кривой и нахождению скорости неравномерного движения. Подобные задачи и задача о вычислении площади, о которой речь будет впереди, интересовали математиков с древних времен. Все же еще в XVI в. постановка таких задач и методы их решения носили глубоко частный характер. Накопившийся в этом направлении обширный материал был приведен в систему и получил теоретическое завершение уже в XVII в. — в работах Ньютона и Лейбница. Очень большой вклад при построении основ современного анализа был внесен Эйлером.

Впрочем Ньютон и Лейбниц и их современники логически мало обосновывали эти великие математические открытия; в способах их рассуждения и в понятиях, которыми они оперировали, с нашей точки зрения можно найти много неясного; да и сами математики того времени это сознавали, о чем свидетельствуют ожесточенные дискуссии, которые происходили по этим вопросам между ними в письмах друг к другу. Математики того времени (XVII—XVIII вв.) особенно тесно связывали свою чисто математическую деятельность с исследовательской деятельностью в различных областях природы (физика, механика, химия, техника). Постановка математической задачи исходила, как правило, из практических потребностей или из желания разобраться в том или ином явлении природы. После того как задача была решена, она так или иначе подвергалась практической проверке, и именно это обстоятельство позволяло целесообразно направлять математические исследования.

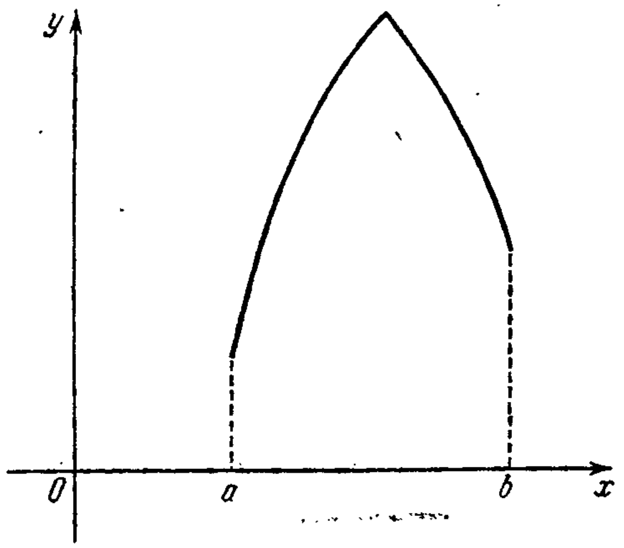


Рис. 12.

Примеры вычисления производных. Определение производной как предела $f'(x)=\lim\limits_{h\to 0}\frac{f(x+h)-f(x)}{h}$ дает возможность для каждой конкретной функции искать ее производную.

Сразу же надо оговорить, что возможны случаи, когда функция в той или иной точке или даже во многих точках просто не имеет производной, т. е. отношение $\frac{f(x+h)-f(x)}{h}$ при $h\to 0$ вовсе не стремится к какому-либо конечному пределу. Этот случай заведомо имеет место во всякой точке разрыва функции $f(x)$, так как при этом в отношении

$$\frac{f(x+h)-f(x)}{h} \tag{10}$$

числитель не стремится к нулю, тогда как знаменатель неограниченно убывает. Производной может не быть и в точке, где функция непрерывна. Простой пример тому дает точка излома (угла) на графике функции (рис. 12). В такой точке кривая графика не имеет определенной касательной, соответственно функция не имеет производной. Часто в таких точках выражение (10) приближается к разным значениям, в зависимости от того, приближается ли h к нулю справа или слева, а если h стремится к нулю произвольным образом, то отношение (10) просто не имеет предела. Пример более сложного случая отсутствия производной дает функция

$$y=\begin{cases} x\sin\frac{1}{x} & \text{при } x\neq 0,\\ 0 & \text{при } x=0.\end{cases}$$

График этой функции приведен на рис. 13. В точке $x = 0$ она не имеет производной, потому что, как видно из графика, в этом случае секущая OA не стремится к определенному положению даже тогда, когда $A \to 0$, приближаясь к 0 с одной стороны. Секущая OA при этом все время колеблется от прямой OM к прямой OL и обратно. Соответственно отношение (10) в этом случае не имеет предела, даже когда h стремится к нулю, сохраняя знак.

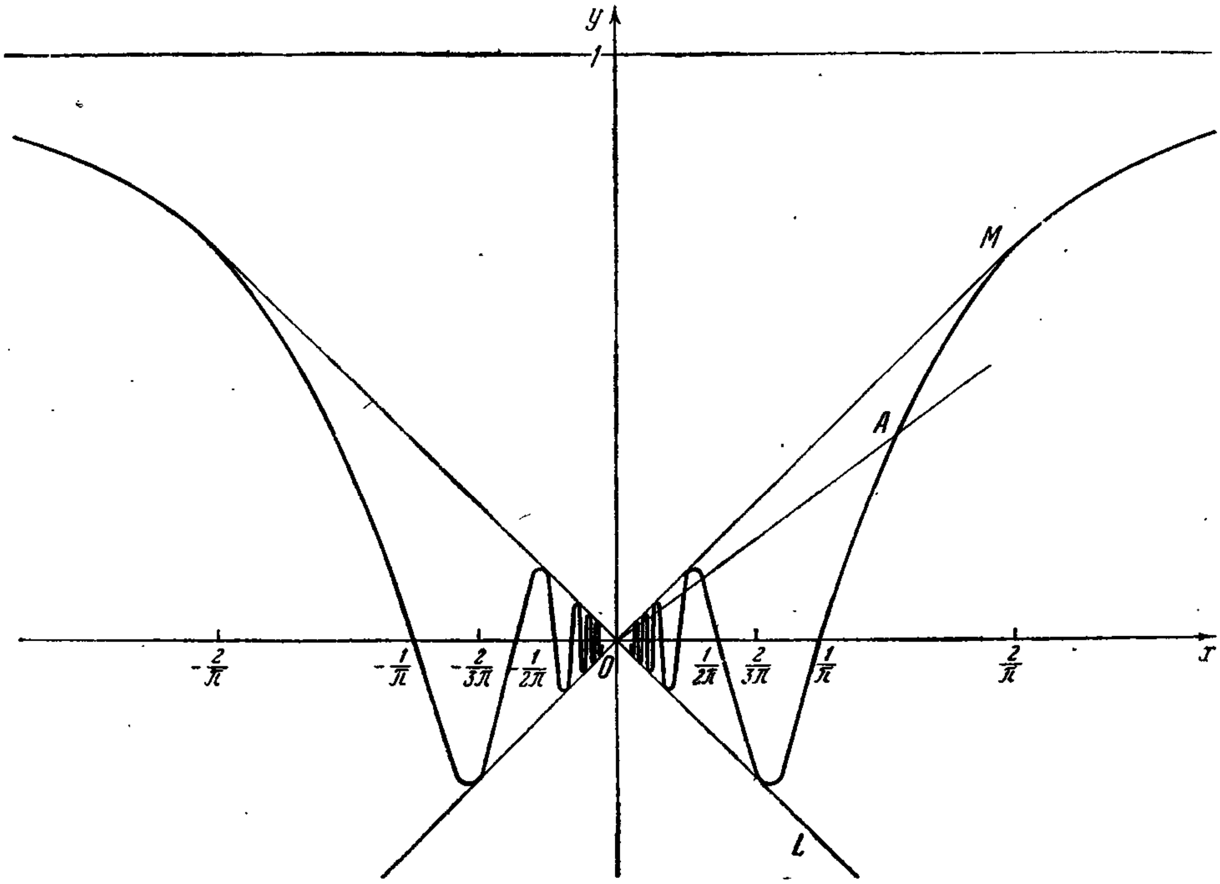


Рис.

Заметим, наконец, что можно задать чисто аналитически, при помощи формулы, функцию непрерывную, но не имеющую ни в одной точке производной. Пример такой функции впервые предложил выдающийся немецкий математик прошлого столетия Вейерштрасс.

Класс дифференцируемых функций, таким образом, значительно у́же класса всех непрерывных функций.

Перейдем к фактическому вычислению производных простейших функций.

1) $y = c$, где c — постоянная. Постоянную можно рассматривать как частный случай функции, которая равна одному и тому же числу для любого x. График ее представляет собой прямую, параллельную оси x-ов, отстоящую от этой оси на расстоянии c. Эта прямая образует с осью x-ов угол $\alpha = 0$, и, очевидно, производная от постоянной тождественно равна нулю: $y' = (c)' = 0$. С точки зрения механики это равенство означает, что скорость неподвижной точки равна нулю.

2) $y = x^2$.

$$\frac{f(x+h)-f(x)}{h} = \frac{(x+h)^2-x^2}{h} = 2x+h.$$

При $h \to 0$ получаем[1] в пределе $2x$, следовательно

$$y' = (x^2)' = 2x.$$

3) $y = x^n$ (n — целое положительное число).

$$\frac{f(x+h)-f(x)}{h} = \frac{(x+h)^n - x^n}{h} = \frac{x^n + nx^{n-1}h + \frac{n(n-1)}{2!}x^{n-2}h^2 + \ldots + h^n - x^n}{h} =$$

$$= nx^{n-1} + \frac{n(n-1)}{2!}x^{n-2}h + \ldots + h^{n-1}.$$

Все слагаемые правой части, начиная со второго, стремятся к нулю при $h \to 0$, поэтому

$$y' = (x^n)' = nx^{n-1}.$$

Эта формула остается верной при любом n, положительном и отрицательном, дробном и даже иррациональном, хотя доказывается это иначе. Мы воспользуемся этим фактом, не приводя его доказательства. Таким образом, например

$$(\sqrt{x})' = \left(x^{\frac{1}{2}}\right)' = \frac{1}{2}x^{-\frac{1}{2}} = \frac{1}{2\sqrt{x}}, \qquad (x>0);$$

$$\left(\sqrt[3]{x}\right)' = \left(x^{\frac{1}{3}}\right)' = \frac{1}{3}x^{-\frac{2}{3}} = \frac{1}{3\sqrt[3]{x^2}}, \qquad (x \neq 0);$$

$$\left(\frac{1}{x}\right)' = (x^{-1})' = -1\cdot x^{-2} = -\frac{1}{x^2}, \quad (x \neq 0).$$

$$(x^{\pi})' = \pi x^{\pi-1}, \qquad (x>0).$$

4) $y = \sin x$.

$$\frac{\sin(x+h)-\sin x}{h} = \frac{2\sin\frac{h}{2}\cos\left(x+\frac{h}{2}\right)}{h} = \frac{\sin\frac{h}{2}}{\frac{h}{2}}\cdot\cos\left(x+\frac{h}{2}\right).$$

Но, как мы выяснили уже, первая дробь при $h \to 0$ стремится к единице, а $\cos\left(x+\frac{h}{2}\right)$, очевидно, к $\cos x$. Итак, производная синуса равна косинусу

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

Предоставляем читателю, рассуждая таким же образом, показать, что

$$(\cos x)' = -\sin x.$$

5) Выше (на стр. 99) уже было отмечено существование предела

$$\lim_{n\to\infty}\left(1+\frac{1}{n}\right)^n = e = 2{,}71828\ldots$$

[1] Мы всюду здесь считаем $h \neq 0$.

Было также указано, что при вычислении этого предела не играет существенной роли то обстоятельство, что n пробегает целые положительные значения. Важно, что бесконечно малая величина $\frac{1}{n}$, добавляемая к единице, и бесконечно растущий по абсолютной величине показатель степени n являются обратными друг другу величинами.

Принимая это утверждение, легко можно найти производную логарифма $y = \log_a x$

$$\frac{\log_a (x+h) - \log_a x}{h} = \frac{1}{h} \log_a \frac{x+h}{x} = \frac{1}{x} \log_a \left(1 + \frac{h}{x}\right)^{\frac{x}{h}}.$$

Непрерывность логарифма позволяет в пределе заменить стоящую под знаком log величину ее пределом, который равен e

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

$\left(\right.$в этом случае роль $n \to \infty$ играет растущая величина $\left.\frac{x}{h}\right)$. В результате мы получаем правило дифференцирования логарифма

$$(\log_a x)' = \frac{1}{x} \log_a e.$$

Это правило становится особенно простым, если за основание логарифмов принято само число e. Логарифм с таким основанием называют *натуральным логарифмом* и обозначают $\ln x$. Можем написать

$$(\log_e x)' = \frac{1}{x},$$

или иначе,

$$(\ln x)' = \frac{1}{x}.$$

§ 6. ПРАВИЛА ДИФФЕРЕНЦИРОВАНИЯ

Из приведенных выше примеров могло показаться, что вычисление производной каждой новой функции требует изобретения новых приемов. Но это не так. Развитию анализа немало способствовало то обстоятельство, что удалось создать весьма простой единый метод, позволяющий находить производную любой «элементарной» функции (функции, которую можно выразить формулой, состоящей из конечной комбинации основных алгебраических действий, тригонометрических функций, возведений в степень и логарифмирований). В основе этого метода лежат так называемые *правила дифференцирования*. Они состоят из ряда теорем, позволяющих сводить более сложные задачи к более простым.

Мы изложим здесь правила дифференцирования, стараясь быть краткими при их выводе. Если читатель намерен почерпнуть из этой главы только самое общее представление об анализе, то он может пропустить этот параграф, запомнив лишь, что есть средство фактически находить производную любой элементарной функции. В этом случае придется, конечно, принять на веру часть вычислений в примерах, которыми будет иллюстрироваться дальнейшее изложение.

Производная суммы. Допустим, что y как функция x задана выражением

$$y = \varphi(x) + \psi(x),$$

где $u = \varphi(x)$ и $v = \psi(x)$ — известные нам функции x. Допустим, кроме того, что производные функций u и v мы находить умеем. Как найти производную функции y? Ответ оказывается простым

$$y' = (u + v)' = u' + v'. \tag{11}$$

В самом деле, дадим x приращение Δx, тогда u, v и y получат в свою очередь приращения Δu, Δv и Δy, связанные равенством

$$\Delta y = \Delta u + \Delta v.$$

Отсюда [1]

$$\frac{\Delta y}{\Delta x} = \frac{\Delta u}{\Delta x} + \frac{\Delta v}{\Delta x},$$

и после перехода к пределу при $\Delta x \to 0$ получим как раз формулу (11), если, конечно, сами функции u и v имели производные.

Аналогично выводится правило дифференцирования разности двух функций

$$(u - v)' = u' - v'. \tag{12}$$

Производная произведения. Несколько сложнее формулируется закон дифференцирования произведения. Производная произведения двух функций, имеющих производные, существует и равна сумме произведений первой функции на производную второй и второй на производную первой, т. е.

$$(uv)' = uv' + vu'. \tag{13}$$

В самом деле, дадим x приращение Δx. Тогда функции u, v и $y = uv$ получат приращения Δu, Δv, Δy, удовлетворяющие соотношению

$$\Delta y = (u + \Delta u)(v + \Delta v) - uv = u\,\Delta v + v\,\Delta u + \Delta u\,\Delta v,$$

откуда

$$\frac{\Delta y}{\Delta x} = u\frac{\Delta v}{\Delta x} + v\frac{\Delta u}{\Delta x} + \Delta u\frac{\Delta v}{\Delta x}.$$

После перехода к пределу при $\Delta x \to 0$ первые два слагаемых правой части дадут правую часть формулы (13), а третье слагаемое исчезнет [2].

---

[1] Здесь всюду $\Delta x \neq 0$.

[2] Последнее слагаемое стремится к нулю при $\Delta x \to 0$, потому что $\frac{\Delta v}{\Delta x}$ стремится к конечному числу, равному производной v', которая с самого начала предполагалась существующей, а $\Delta u \to 0$ (поскольку функция u, как имеющая по условию производную, непрерывна).

В пределе из последнего равенства получим правило (13).

В частном случае, если $v = c = \text{const}$, то

$$(cu)' = cu' + uc' = cu', \tag{14}$$

так как производная постоянной равна нулю.

Производная частного. Пусть $y = \frac{u}{v}$, где u и v — функции, имеющие при данном x производную, причем $v \neq 0$ для этого x. Очевидно

$$\Delta y = \frac{u + \Delta u}{v + \Delta v} - \frac{u}{v} = \frac{v\,\Delta u - u\,\Delta v}{(v + \Delta v)\,v},$$

откуда

$$\frac{\Delta y}{\Delta x} = \frac{v\frac{\Delta u}{\Delta x} - u\frac{\Delta v}{\Delta x}}{(v + \Delta v)\,v} \to \frac{vu' - uv'}{v^2} \quad (\Delta x \to 0).$$

Здесь мы снова воспользовались тем, что для функции v, имеющей производную, обязательно $\Delta v \to 0$, когда $\Delta x \to 0$. Таким образом,

$$\left(\frac{u}{v}\right)' = \frac{vu' - uv'}{v^2}. \tag{15}$$

Вот несколько примеров на применение полученных правил:

$$(2x^3 - 5)' = 2(x^3)' - (5)' = 2 \cdot 3x^2 - 0 = 6x^2;$$

$$(x^2 \sin x)' = x^2 (\sin x)' + (x^2)' \sin x = x^2 \cos x + 2x \sin x;$$

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

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

Предоставим читателю самому доказать формулу:

$$(\operatorname{ctg} x)' = -\operatorname{cosec}^2 x.$$

Производная обратной функции. Рассмотрим функцию $y = f(x)$, непрерывную и возрастающую (убывающую) на отрезке $[a, b]$. Последнее означает, что большему значению x из отрезка $[a, b]$ соответствует большее (меньшее) значение y (рис. 14).

Пусть $c = f(a)$ и $d = f(b)$. На рис. 14 видно, что каждому значению y из отрезка $[c, d]$ (соотв. $[d, c]$) соответствует, и притом одно, значение x из отрезка $[a, b]$ такое, что $y = f(x)$. Этим мы задали на отрезке $[c, d]$ (соотв. $[d, c]$) вполне определенную функцию $x = \varphi(y)$, которая называется *обратной функцией* по отношению к функции $y = f(x)$. На рис. 14 видно, что функция $\varphi(y)$ непрерывна. Впрочем, этот факт в современном анализе строго доказывается на аналитической основе. Пусть теперь Δx и Δy — соответствующие друг другу приращения x и y. Очевидно

$$\frac{\Delta y}{\Delta x} = \frac{1}{\frac{\Delta x}{\Delta y}}, \text{ если } \Delta y \neq 0.$$

В пределе это дает нам простое соотношение для производных прямой и обратной функции

$$y'_x = \frac{1}{x'_y}. \tag{16}$$

Используем это соотношение для нахождения производной функции $y = a^x$. Обратную функцию $x = \log_a y$ мы умеем дифференцировать, что позволяет написать

$$(a^x)'_x = \frac{1}{(\log_a y)'_y} = \frac{1}{\frac{1}{y}\log_a e} = y \log_e a = a^x \ln a. \tag{17}$$

В частности $(e^x)' = e^x$.

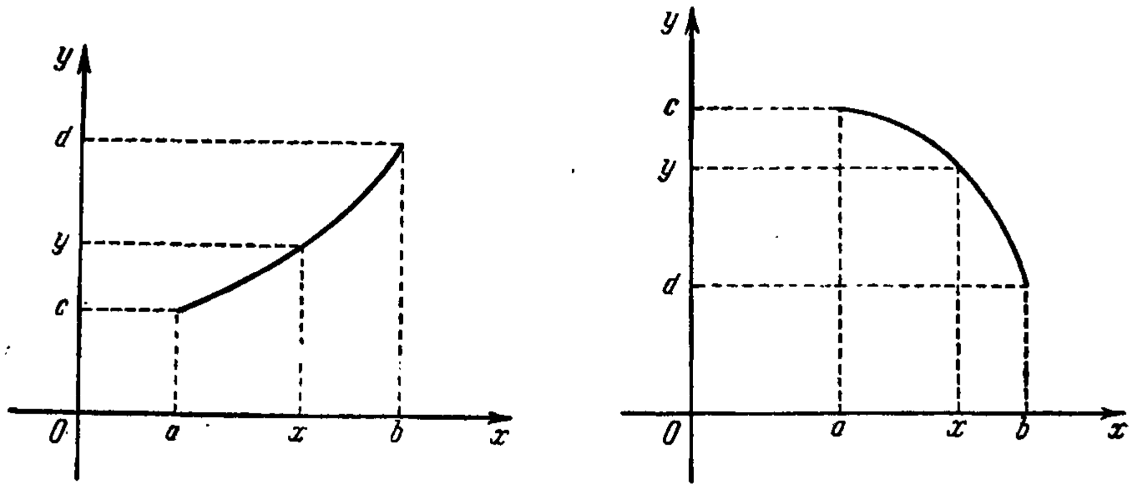


Рис. 14.

Другой пример: $y = \arcsin x$. Обратная функция $x = \sin y$. Поэтому

$$(\arcsin x)'_x = \frac{1}{(\sin y)'_y} = \frac{1}{\cos y} = \frac{1}{\sqrt{1-(\sin y)^2}} = \frac{1}{\sqrt{1-x^2}}.$$

Таблица производных. Перечислим производные простейших элементарных функций:

y	y'	y	y'	y	y'
c	0	$\ln x$	$\frac{1}{x}$	$\operatorname{tg} x$	$\sec^2 x$
x^a	ax^{a-1}	$\log_a x$	$\frac{1}{x}\log_a e$	$\arcsin x$	$\frac{1}{\sqrt{1-x^2}}$
e^x	e^x	$\sin x$	$\cos x$	$\arccos x$	$-\frac{1}{\sqrt{1-x^2}}$
a^x	$a^x \ln a$	$\cos x$	$-\sin x$	$\operatorname{arctg} x$	$\frac{1}{1+x^2}$

Эти формулы были нами выведены и разъяснены, за исключением двух последних, которые читатель, если пожелает, может вывести сам, воспользовавшись правилом дифференцирования обратной функции.

Нахождение производной функции от функции. Нам осталось рассмотреть последнее, наиболее трудное правило дифференцирования. Владеющий этим правилом и таблицей может с полным основанием считать, что он сумеет продифференцировать любую элементарную функцию.

Чтобы применять правило, которое мы здесь дадим, нужно совершенно четко представлять себе, как устроена функция, которую требуется продифференцировать, какие действия (операции) и в каком порядке надо, исходя от независимой переменной x, произвести, чтобы, в конце концов, получить функцию y.

Например, чтобы вычислить функцию

$$y = \sin x^2,$$

надо сначала возвести x в квадрат, а затем от полученной величины найти синус. Это можно записать следующим образом: $y = \sin u$, где $u = x^2$.

Напротив, для вычисления функции

$$y = \sin^2 x$$

надо сначала найти синус от x, а затем полученное значение возвести в квадрат, что можно записать так: $y = u^2$, где $u = \sin x$.

Вот еще примеры:

1) $y = (3x+4)^3$, $\quad y = u^3$, $\quad u = 3x+4$.

2) $y = \sqrt{1-x^2}$, $\quad y = u^{1/2}$, $\quad u = 1-x^2$.

3) $y = e^{kx}$; $\quad y = e^u$, $\quad u = kx$.

В более сложных случаях получается цепочка простых зависимостей, имеющая несколько звеньев. Например

4) $y = \cos^3 x^2$; $\quad y = u^3$; $\quad u = \cos v$; $\quad v = x^2$.

Если y есть функция от переменной u

$$y = f(u), \tag{18}$$

а u в свою очередь есть функция от переменной x

$$u = \varphi(x), \tag{19}$$

то y, будучи функцией от u, есть некоторая функция от x, которую мы обозначим так:

$$y = F(x) = f[\varphi(x)]. \tag{20}$$

Усложняя этот процесс, можно образовать, например, функцию

$$y = \Phi(x) = f\{\varphi[\psi(x)]\},$$

которая эквивалентна равенствам

$$y = f(u), \quad u = \varphi(v), \quad v = \psi(x),$$

или получить еще более сложную функцию, сводящуюся к цепочке подобных равенств.

Мы сейчас покажем, как вычислить производную от функции $F(x)$, определяемой равенством (20), если известны производная от $f(u)$ по u и производная от $\varphi(x)$ по x.

Дадим x приращение Δx, тогда u получит в силу равенства (19) некоторое приращение Δu, а y в силу равенства (18) — приращение Δy. Можно написать

$$\frac{\Delta y}{\Delta x}=\frac{\Delta y}{\Delta u}\cdot\frac{\Delta u}{\Delta x}.$$

Пусть теперь Δx стремится к нулю. При этом $\frac{\Delta u}{\Delta x}\to u'_x$. Далее, вследствие непрерывности u, приращение $\Delta u\to 0$, и потому $\frac{\Delta y}{\Delta u}\to y'_u$ (существование производных y'_u и u'_x было предположено).

Этим доказана важная формула производной функции от функции[1]

$$y'_x=y'_u u'_x. \tag{21}$$

Вычислим, воспользовавшись формулой (21) и основной таблицей производных (см. стр. 114), производные от функций, приведенных выше в качестве примеров:

1) $y=(3x+4)^3=u^3,\qquad y'_x=(u^3)'_u(3x+4)'_x=3u^2\cdot 3=9(3x+4)^2.$

2) $y=\sqrt{1-x^2}=u^{\frac{1}{2}},\quad y'_x=\left(u^{\frac{1}{2}}\right)'_u(1-x^2)'_x=\frac{1}{2}u^{-\frac{1}{2}}(-2x)=-\frac{x}{\sqrt{1-x^2}}.$

3) $y=e^{kx}=e^u,\qquad y'_x=(e^u)'_u\cdot u'_x=e^u\cdot k=ke^{kx}.$

Если $y=f(u),\quad u=\varphi(v),\quad v=\psi(x),$

то

$$y'_x=y'_u\cdot u'_x=y'_u(u'_v\cdot v'_x)=y'_u\cdot u'_v\cdot v'_x.$$

Ясно, как обобщается эта формула в случае любого (конечного) числа входящих в цепочку функций.

Пример.

4) $y=\cos^3 x^2;\qquad y'_x=(u^3)'_u(\cos v)'_v\cdot(x^2)'_x=3u^2(-\sin v)\cdot 2x=$
$=-6x\cos^2 x^2\sin x^2.$

Чтобы пояснить, как вычисляются производные функции от функций, мы вводили промежуточные переменные u, v,... На самом деле, после некоторой тренировки можно освободиться от этого, держа эти обозначения в уме.

Элементарные функции. Заканчивая этот параграф, отметим, что функции, список производных которых был приведен выше в виде табли-

---

[1] При выводе этой формулы мы неявно предполагали, что при стремлении Δx к нулю Δu все время не равно нулю. На самом деле она остается верной и тогда, когда это предположение не имеет места.

цы, можно положить в основу определения так называемых элементарных функций. Именно все функции, которые получаются из этих простейших функций с помощью четырех арифметических действий и операций функции от функции, примененных конечное число раз, называются *элементарными*.

Например, многочлен $x^3 - 2x^2 + 3x - 5$ есть элементарная функция, так как она получается из нескольких функций вида x^k при помощи арифметических операций. Функция $\ln\sqrt{1-x^2}$ тоже элементарная, так как она получается из многочлена $u = 1 - x^2$ при помощи операции $v = u^{1/2}$, а затем операции $\ln v$.

Изложенных выше правил дифференцирования достаточно, чтобы найти производную любой элементарной функции, зная производные простейших элементарных функций.

§ 7. МАКСИМУМ И МИНИМУМ. ИССЛЕДОВАНИЕ ГРАФИКОВ ФУНКЦИЙ

Одним из простейших и важных приложений производной является теория максимума и минимума. Допустим, что на некотором участке $a \leqslant x \leqslant b$ задана функция $y = f(x)$, о которой мы будем предполагать, что

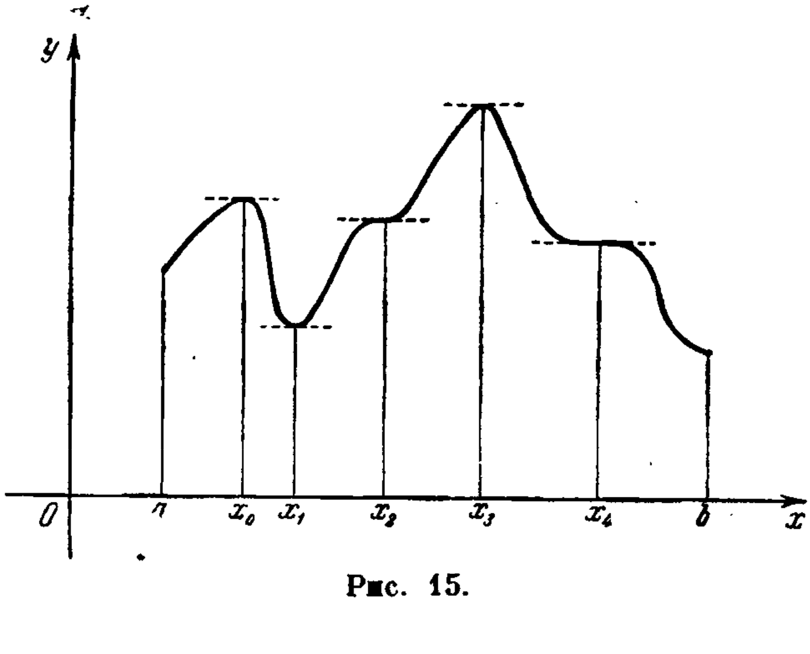


Рис. 15.

она не только непрерывна, но и имеет во всех точках производную. Умение вычислять производную дает возможность ясно представить себе ход графика функции. На участке, где производная все время положительна, касательная к графику идет вверх. Функция на таком участке возрастает, т. е. большему значению x соответствует большее значение $f(x)$. На участке, где производная все время отрицательна, функция, наоборот, убывает; график идет вниз.

Максимум и минимум. На рис. 15 изображен график функции $y = f(x)$, определенной на отрезке $[a, b]$. Особенный интерес представляют точки этого графика, имеющие абсциссы x_0, x_1, x_3.

Говорят, что функция $f(x)$ имеет в точке x_0 местный максимум; этим хотят выразить, что в точке x_0 функция $f(x)$ больше, чем в соседних точках, точнее, $f(x_0) \geqslant f(x)$ для всех x из некоторого отрезка, окружающего точку x_0.

Аналогично определяется местный минимум.

Для нашей функции местный максимум достигается в точках x_0 и x_3, а местный минимум — в точке x_1.

В каждой точке минимума или максимума, если она есть внутренняя точка отрезка $[a, b]$, т. е. не совпадающая с его концами a и b, производная должна равняться нулю.

Последнее, весьма важное утверждение следует из самого определения производной как предела отношения $\frac{\Delta y}{\Delta x}$. В самом деле, при небольшом сдвиге из точки максимума $\Delta y \leqslant 0$. Поэтому при положительных Δx отношение $\frac{\Delta y}{\Delta x}$ неположительно, а при отрицательных Δx отношение $\frac{\Delta y}{\Delta x}$ неотрицательно. Предел этого отношения, который по предположению существует, уже не может быть ни положительным, ни отрицательным, и ему остается быть только нулем. Наглядно это соответствует тому, что в точках максимума или минимума (обычно опускают слово «местный», хотя и подразумевают его) касательная к графику горизонтальна. На рис. 15 можно заметить, что в точках x_2 и x_4 касательная тоже горизонтальна, как и в точках x_0, x_1, x_3, хотя в этих точках нет ни максимума, ни минимума. Точек, в которых производная функции равна нулю (*стационарных точек*), может, вообще говоря, оказаться больше, чем точек максимума или минимума.

Отыскание наибольших и наименьших значений функции. В самых разнообразных технических вопросах бывает необходимо узнать, при каком x та или иная функция $f(x)$ достигает наибольшего или наименьшего значения по отношению к данному отрезку.

В случае, когда надо найти наибольшее значение функции, речь идет об отыскании на отрезке $[a, b]$ точки x_0, для которой при всех x из $[a, b]$ выполняется неравенство $f(x_0) \geqslant f(x)$.

Однако возникает принципиальный вопрос, существует ли вообще хотя бы одна такая точка. Средствами современного анализа можно доказать теорему существования: если функция $f(x)$ непрерывна на конечном отрезке, то на нем существует по меньшей мере одна точка, в которой функция достигает максимума (минимума) по отношению к отрезку $[a, b]$.

Из сказанного выше вытекает, что точки максимума и минимума следует искать прежде всего среди «стационарных» точек. На этом основывается следующий известный способ отыскания максимума и минимума.

Находим производную от $f(x)$ и, приравняв ее нулю, решаем полученное уравнение

$$f'(x) = 0.$$

Если $x_1, x_2, \ldots, x_n$—корни этого уравнения, то сравниваем между собой числа $f(x_1), f(x_2), \ldots, f(x_n)$. Конечно, надо еще учесть, что максимум и минимум может оказаться не внутри, а на краю промежутка $[a, b]$ (как для минимума на рис. 15) или в точке, где функция не имеет производной (как на рис. 12). Поэтому к точкам $x_1, x_2, \ldots, x_n$ добавляют еще концы a и b промежутка и точки, в которых нет производной, если такие точки есть. Остается, сравнив значения функции во всех этих точках, выбрать среди них наибольшее или наименьшее.

По поводу высказанной выше теоремы существования важно добавить, что она уже, вообще говоря, перестает быть верной в случае, если функция $f(x)$ непрерывна только на и н т е р в а л е (a, b), т. е. на множестве точек x, удовлетворяющих неравенствам $a < x < b$. Предоставляем читателю разобраться в том, что функция $\frac{1}{x}$ на интервале $(0, 1)$ не имеет ни максимума, ни минимума.

Разберем примеры.

Из квадратного куска жести со стороной a требуется сделать прямоугольную открытую коробку наибольшего объема. Если мы из углов куска вырежем квадратики со стороной x (см. § 2, пример 2), то получим коробку объемом

$$V = x(a - 2x)^2.$$

Наша задача сводится к разысканию x, при котором функция $V(x)$ достигает наибольшего значения на отрезке $0 \leqslant x \leqslant \frac{a}{2}$. Согласно правилу, найдем производную и приравняем ее нулю

$$V'(x) = (a - 2x)^2 - 4x(a - 2x) = 0.$$

Решая полученное уравнение, найдем два корня

$$x_1 = \frac{a}{2}, \qquad x_2 = \frac{a}{6}.$$

К ним еще присоединим левый конец отрезка, на котором задана функция $V(x)$ (правый конец совпал с x_1). Сравним значения функции в этих точках

$$V(0) = 0; \quad V\left(\frac{a}{6}\right) = \frac{2}{27}a^3, \qquad V\left(\frac{a}{2}\right) = 0.$$

Итак, коробка будет иметь наибольший объем, равный $\frac{2}{27}a^3$, при высоте $x = \frac{a}{6}$.

В качестве второго примера разберем задачу с фонарем (см. § 2, пример 3). На какой высоте h надо повесить фонарь, чтобы конькобежная дорожка была максимально освещена?

В силу формулы (3) § 2 наша задача сводится к определению h, при котором $T = \frac{A \sin \alpha}{h^2 + r^2}$ принимает наибольшее значение. Вместо h удобнее искать угол α (рис. 3 на стр. 87). Имеем

$$h = r \operatorname{tg} \alpha,$$

следовательно

$$T = \frac{A}{r^2}\frac{\sin\alpha}{1+\operatorname{tg}^2\alpha} = \frac{A}{r^2}\sin\alpha\cos^2\alpha.$$

Требуется найти максимум функции $T(\alpha)$ среди значений α, удовлетворяющих неравенству $0<\alpha<\frac{\pi}{2}$. Найдем производную и приравняем ее нулю

$$T'(\alpha) = \frac{A}{r^2}(\cos^3\alpha - 2\sin^2\alpha\cos\alpha) = 0.$$

Это уравнение распадается на два

$$\cos\alpha = 0, \qquad \cos^2\alpha - 2\sin^2\alpha = 0.$$

Первое уравнение дает корень $\alpha = \frac{\pi}{2}$, совпадающий с концом промежутка $\left(0, \frac{\pi}{2}\right)$. Второму уравнению можно придать вид

$$\operatorname{tg}^2\alpha = \frac{1}{2}.$$

Но так как $0<\alpha<\frac{\pi}{2}$, то $\alpha\approx 35°15'$. Это и есть значение, при котором функция $T(\alpha)$ достигает максимума (на концах промежутка функция меньше, так как там $T=0$). Искомая высота h будет равна

$$h = r\operatorname{tg}\alpha = \frac{r}{\sqrt{2}} \approx 0{,}7r.$$

Для наилучшего освещения льда фонарь должен быть поднят на высоту около $0{,}7\,r$.

Допустим теперь, что имеющееся устройство не позволяет поднять фонарь на высоту, большую некоторого H. Тогда угол α может изменяться не от 0 до $\frac{\pi}{2}$, а в более узких пределах: $0<\alpha\leqslant\operatorname{arctg}\frac{H}{r}$. Пусть, например, $r=12$ *м*, $H=9$ *м*. В этом случае фонарь действительно можно установить на высоте $h=\frac{r}{\sqrt{2}}$, составляющей несколько более 8 *м*, что и надо сделать. Но если H меньше 8 *м* (например, если для установки фонаря в нашем распоряжении есть столб длиной всего 6 *м*), то окажется, что функция $T(\alpha)$ в промежутке $\left[0, \operatorname{arctg}\frac{H}{r}\right]$ не имеет равной нулю производной. В этом случае наибольшее значение достигается на краю промежутка, и фонарь следует поднять до наибольшей имеющейся высоты $H=6$ *м*.

До сих пор мы рассматривали функцию на конечном отрезке. Если участок бесконечен, то даже непрерывная функция может не достигать в нем наибольшего или наименьшего значения, а все время, например, возрастать или убывать при стремлении x к бесконечности.

Так, например, функции $y=kx+b$ (см. рис. 5, стр. 89), $y=\operatorname{arctg}x$ (рис. 16а), $y=\ln x$ (рис. 16б) нигде не достигают ни максимума, ни минимума. Функция $y=e^{-x^2}$ (рис. 16в) достигает максимума в точке $x=1$, но нигде не достигает минимума. Что же касается функции $y = \frac{x}{1+x^2}$ (рис. 16г), то она достигает минимума в точке $x = -1$ и максимума в точке $x = 1$.

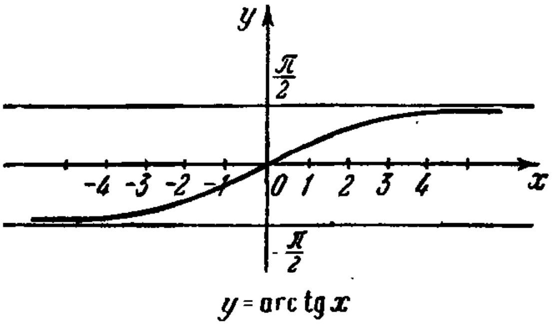


Рис. 16 а.

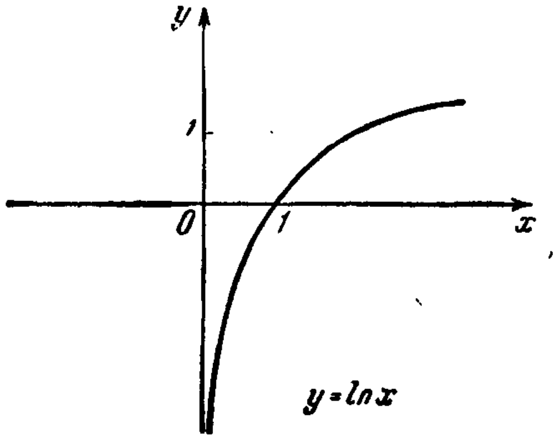


Рис. 16 б.

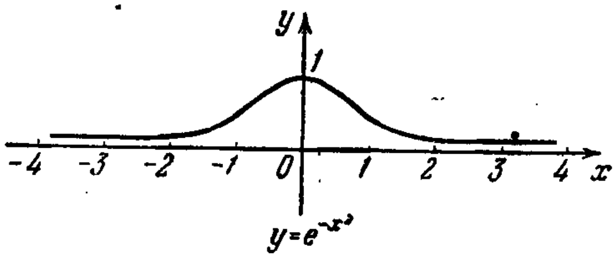


Рис. 16 в.

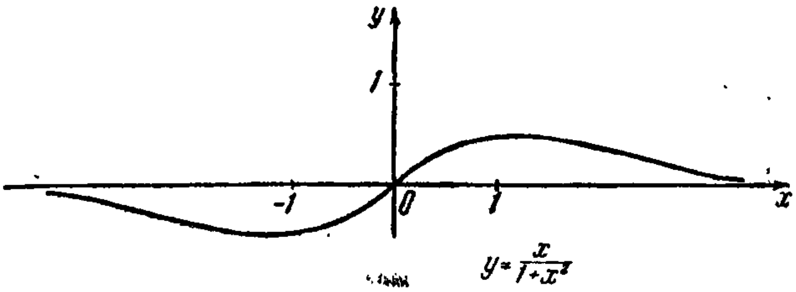


Рис. 16 г.

В случае бесконечного промежутка исследование ведется по обычным правилам. Только вместо $f(a)$ и $f(b)$ следует рассматривать пределы

$$A = \lim_{x \to -\infty} f(x), \qquad B = \lim_{x \to +\infty} f(x).$$

Производные высших порядков. Уже для того, чтобы более детально исследовать графики функций, нам придется изучить ход изменения производной $f'(x)$ рассматриваемой функции $f(x)$. Выражение $f'(x)$ в свою очередь является некоторой функцией от x. От нее также можно найти производную.

Производную от производной называют второй производной и обозначают

$$[y']' = y'' \quad \text{или} \quad [f'(x)]' = f''(x).$$

Аналогично можно вычислить третью производную

$$[y'']' = y''' \text{ или } [f''(x)]' = f'''(x)$$

и т. д., вообще n-ю производную или, как еще говорят, производную n-го порядка

$$y^{(n)} = f^{(n)}(x).$$

Надо, конечно, иметь в виду, что эта цепь может для какого-либо x (или даже для всех x) оборваться на некоторой k-й производной: может случиться, что $f^{(k)}(x)$ существует, а производная $f^{(k+1)}(x)$ уже не будет существовать. Производные любых порядков будут применяться нами ниже, в § 9, при рассмотрении формулы Тейлора. Сейчас мы остановимся на второй производной.

Смысл второй производной. Выпуклость и вогнутость. Вторая производная имеет простой механический смысл. Пусть $s = f(t)$ есть закон прямолинейного движения точки, тогда s' — скорость, а s'' — «скорость изменения скорости» или, проще, ускорение точки в момент времени t. Например, для случая падения тел

$$s = \frac{gt^2}{2} + v_0 t + s_0,$$

$$s' = gt + v_0,$$

$$s'' = g,$$

т. е. ускорение падающих тел постоянно.

Вторая производная имеет простой геометрический смысл. Как по знаку первой производной можно определить, возрастает или убывает функция,

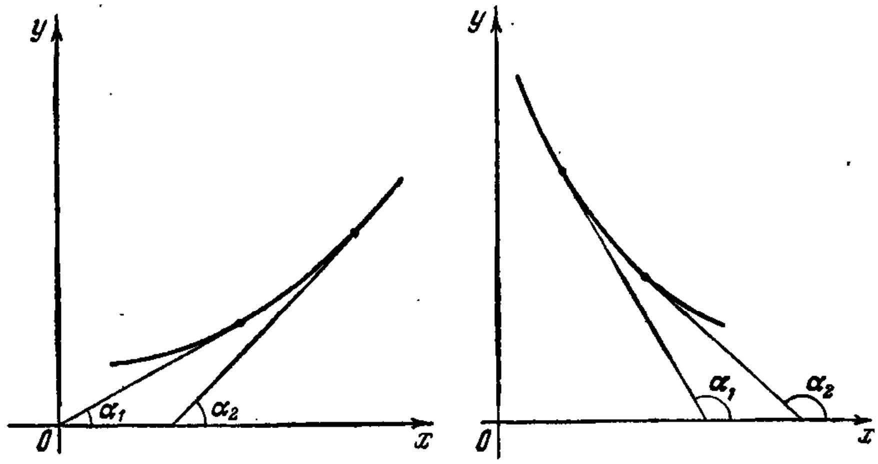


Рис. 17.

так по знаку второй производной можно судить, в какую сторону изгибается линия графика функции.

Если на некотором участке вторая производная все время положительна, то первая производная возрастает, следовательно, возрастает $f'(x) = \operatorname{tg} \alpha$, а потому растет и сам угол α наклона касательной к графику (рис. 17). При этом по мере продвижения вдоль кривой с переходом к последующим участкам кривая поворачивается все время в одну и ту же сторону и располагается, как говорят, «выпуклостью вниз».

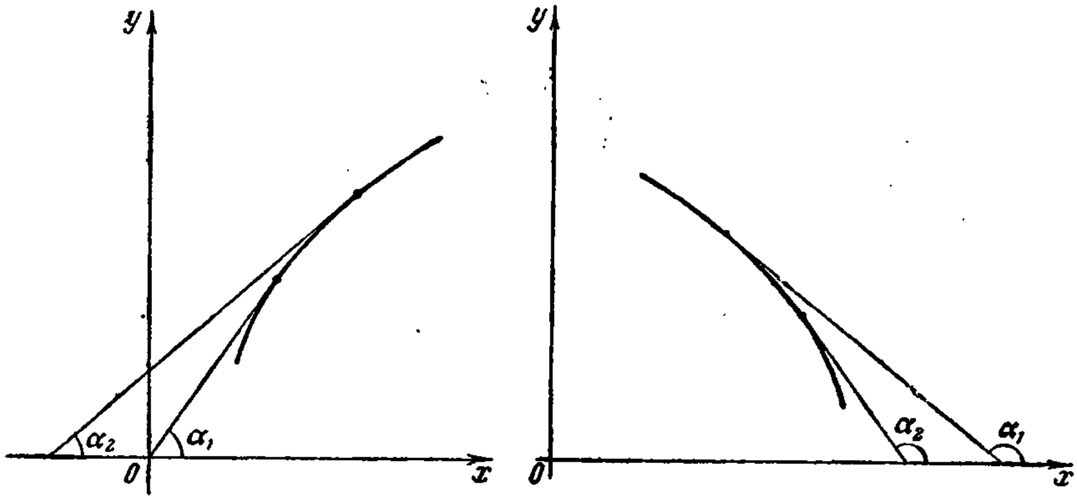


Рис. 18.

Напротив, на участке, где вторая производная все время отрицательна (рис. 18), линия графика функции оказывается изогнутой «выпуклостью вверх» [1].

Признаки максимумов и минимумов. Исследование графиков функций. Если кривая на протяжении всего данного отрезка изменения x обращена выпуклостью кверху и в некоторой точке x_0 этого отрезка имеет производную, равную нулю, то в этой точке она необходимо достигает максимума; в случае же выпуклости книзу — минимума. Это простое соображение позволяет часто, обнаружив точку, в которой производная равна нулю, выяснить затем, имеется ли в этой точке местный максимум или минимум [2].

Пример 1. Исследуем, как выглядит график функции

$$f(x) = \frac{x^3}{3} - \frac{5x^2}{2} + 6x - 2.$$

Возьмем ее первую производную и приравняем нулю

$$f'(x) = x^2 - 5x + 6 = 0.$$

Корнями полученного уравнения будут: $x_1 = 2$, $x_2 = 3$. Соответствующие значения функции

$$f(2) = 2^2/_3, \quad f(3) = 2^1/_2.$$

---

[1] Строгим определением «выпуклости вверх» является свойство кривой располагаться выше (точнее сказать: «не ниже») хорды, соединяющей любые две ее точки; аналогично при «выпуклости вниз» или, как кратко говорят, «вогнутости» кривая проходит не выше своих хорд.

[2] В более сложных случаях, когда вторая производная сама меняет знак, задача о выяснении характера стационарной точки решается при помощи формулы Тейлора (§ 9).

Отметим полученные две точки на чертеже. К ним еще можно присоединить точку с координатами $x=0$ и $y=f(0)=-2$, в которой график пересекает ось y-ов. Вторая производная равна $f''(x)=2x-5$. Она обращается в нуль при $x=\frac{5}{2}$, причем

$$f''(x)>0 \text{ при } x>{}^5/_2,$$

$$f''(x)<0 \text{ при } x<{}^5/_2.$$

Точка

$$x=\frac{5}{2}, \quad y=f\left(\frac{5}{2}\right)=2^7/_{12}$$

есть *точка перегиба* графика. Слева от нее кривая выпуклая кверху, справа — выпуклая книзу.

Очевидно теперь, что точка $x=2$ есть точка максимума, а точка $x=3$ — точка минимума функции.

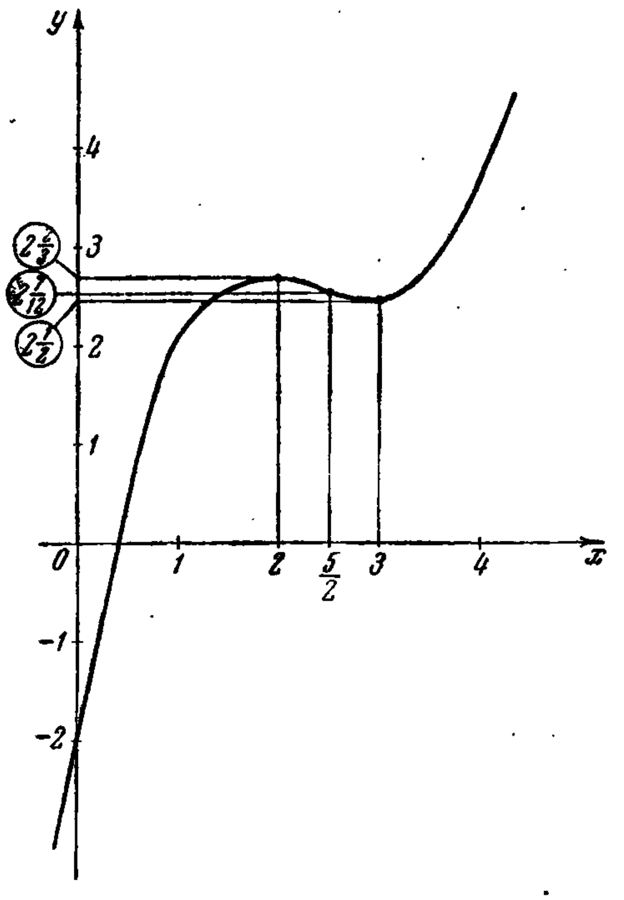


Рис. 19.

На основании полученных данных заключаем, что график функции $y=f(x)$ имеет вид, изображенный на рис. 19. Кривая от точки $(0, -2)$ с возрастанием x возрастает, будучи обращенной выпуклостью кверху, достигает в точке $(2, 2^2/_3)$ своего максимума, затем опускается вниз. В точке $(2^1/_2, 2^7/_{12})$, где $f''(x)=0$, выпуклость меняется на вогнутость. Далее, в точке $(3, 2^1/_2)$ достигается минимум функции, а затем при дальнейшем возрастании x кривая возрастает до бесконечности. Последнее утверждение вытекает из того, что первый член функции, содержащий наивысшую (третью) степень x, стремится к бесконечности быстрее, чем второй и третий. На этом же основании график функции уходит к $-\infty$, когда x, принимая отрицательные значения, возрастает по абсолютной величине.

Пример 2. Покажем, что для любого x имеет место неравенство $e^x \geqslant 1+x$. Для этого рассмотрим функцию $f(x)=e^x-x-1$. Ее первая производная равна $f'(x)=e^x-1$ и обращается в нуль только при $x=0$. Вторая производная $f''(x)=e^x>0$ для всех x. Следовательно, график функции

$f(x)$ — выпуклый книзу. Число $f(0)=0$ есть минимум нашей функции, и $e^x - x - 1 \geqslant 0$ для всех x.

Исследование графиков может преследовать самые различные цели. С их помощью, например, часто выясняют число действительных корней того или иного уравнения. Так, чтобы доказать, что уравнение

$$xe^x = 2$$

имеет единственный действительный корень, можно исследовать графики функций $y = e^x$ и $y = \frac{2}{x}$ (они изображены на рис. 20). Легко видеть, что графики этих функций пересекаются только в одной точке, и, следовательно, уравнение $e^x = \frac{2}{x}$ имеет единственный корень.

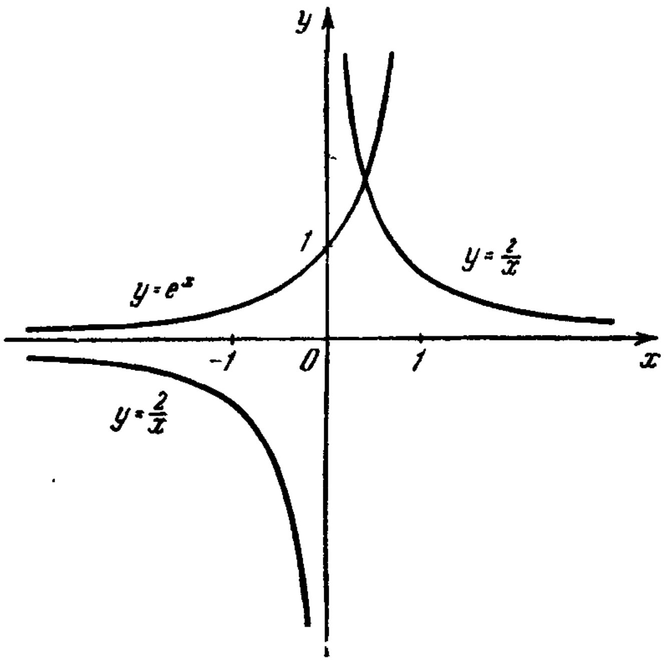


Рис. 20.

Методы анализа широко применяются к вопросам приближенного вычисления корней уравнений. По этому поводу см. главу IV, § 5.

§ 8. ПРИРАЩЕНИЕ И ДИФФЕРЕНЦИАЛ ФУНКЦИИ

Дифференциал функции. Рассмотрим функцию $y = f(x)$, имеющую производную. Приращение этой функции

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

соответствующее приращению аргумента Δx, обладает тем свойством, что отношение $\frac{\Delta y}{\Delta x}$ при $\Delta x \to 0$ стремится к конечному пределу, равному производной

$$\frac{\Delta y}{\Delta x} \to f'(x).$$

То же самое можно записать в виде равенства

$$\frac{\Delta y}{\Delta x} = f'(x) + \alpha,$$

где α — величина, зависящая от Δx, и притом такая, что при $\Delta x \to 0$, она тоже стремится к нулю. Отсюда приращение функции представляется в виде

$$\Delta y = f'(x)\,\Delta x + \alpha\,\Delta x,$$

где $\alpha \to 0$, если $\Delta x \to 0$.

Первое слагаемое правой части этого равенства весьма просто зависит от Δx, именно: оно пропорционально Δx. Его называют *дифференциалом* функции для заданного значения x, соответствующим данному приращению аргумента Δx, и обозначают

$$dy = f'(x)\,\Delta x.$$

Второе слагаемое характерно тем, что оно стремится к нулю при $\Delta x \to 0$ быстрее, чем Δx, благодаря наличию в нем множителя α. Говорят, что второе слагаемое есть бесконечно малая высшего порядка по отношению к Δx, а в случае, если $f'(x) \neq 0$,— и по отношению к первому слагаемому. Этим хотят сказать, что при достаточно малом Δx второе слагаемое не только само мало, но и его отношение к Δx становится сколь угодно малым.

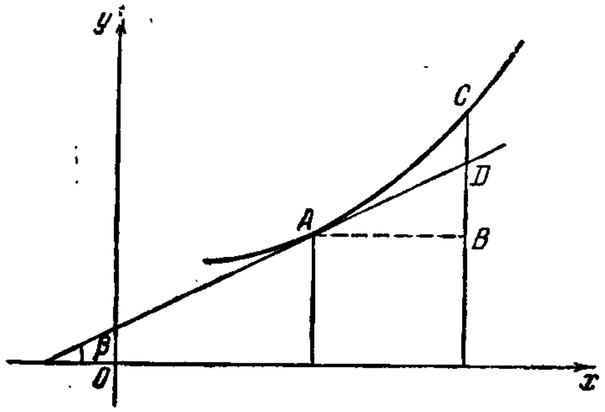


Рис. 21.

Это разложение Δy на два слагаемых, из которых первое (главная часть) линейно зависит от Δx, а второе ничтожно мало при малых Δx, можно проследить на рис. 21. Отрезок $BC = \Delta y$, при этом $BC = BD + DC$, где $BD = \operatorname{tg}\beta \cdot \Delta x = f'(x)\,\Delta x = dy$, а DC — бесконечно малая высшего порядка относительно Δx.

На практике часто пользуются дифференциалом для приближенного представления приращения функции. Например, пусть требуется определить объем стенок закрытой кубической коробки, внутренние линейные размеры которой равны $10 \times 10 \times 10$ см, а толщина стенок 0,05 см. Если не требуется при этом особой точности, можно рассуждать так. Объем всех стенок коробки представляет собой приращение Δy функции $y = x^3$ при $x = 10$, соответствующее $\Delta x = 0{,}1$. Приближенно мы находим:

$$\Delta y \approx dy = (x^3)'\,\Delta x = 3x^2\,\Delta x = 3 \cdot 10^2 \cdot 0{,}1 = 30 \text{ см}^3.$$

В целях симметрии обозначений приращение Δx независимой переменной принято обозначать символом dx и также называть его дифференциалом. При таком обозначении дифференциал функции будет записываться так:

$$dy = f'(x)\,dx.$$

Отсюда производная есть отношение $f'(x) = \frac{dy}{dx}$ дифференциала функции к дифференциалу независимого переменного.

Дифференциал функции исторически ведет свое происхождение от понятия «неделимой». Это с современной точки зрения далеко не четкое понятие в свое время, в XVII в., было основным в математическом анализе. Представление об этом понятии в течение нескольких столетий претерпело существенные изменения. Неделимую, а затем дифференциал функции, представляли как актуальную бесконечно малую величину — что-то вроде очень малой постоянной величины, не являющейся в то же время нулем. Выше было дано определение дифференциала, как оно понимается в современном анализе. В силу этого определения дифференциал есть величина конечная для каждого приращения аргумента Δx и притом пропорциональная ему. Другое основное свойство дифференциала — характер отличия его от Δy — можно познать лишь в движении, именно: если мы будем рассматривать стремящееся к нулю (бесконечно малое) приращение Δx, то разница между dy и Δy будет при этом становиться сколь угодно малой даже по отношению к Δx.

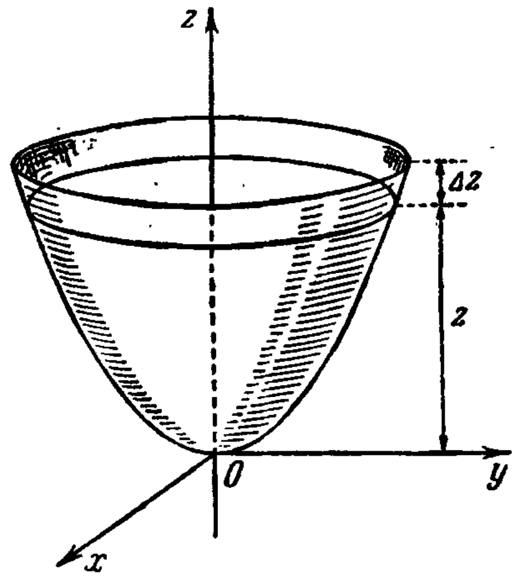


Рис. 22.

На замене малых приращений дифференциалами строится большинство приложений анализа бесконечно малых к исследованию явлений природы. Особенно ясно читатель увидит это на примере дифференциальных уравнений, которым в этой книге посвящены главы V и VI (том 2).

Чтобы узнать функцию, выражающую данный процесс, стараются получить сначала уравнение, связывающее определенным образом эту функцию с ее производными того или иного порядка. Метод получения такого уравнения, называемого дифференциальным, часто сводят к замене приращений искомых функций соответствующими дифференциалами.

Решим для примера следующую задачу. В пространстве, где задана прямоугольная система координат $Oxyz$, рассмотрим поверхность, полученную вращением параболы, уравнение которой (в плоскости Oyz) имеет вид $z = y^2$. Эта поверхность называется параболоидом вращения (рис. 22). Пусть v обозначает объем тела, ограниченного параболоидом и плоскостью, параллельной плоскости Oxy и отстоящей от нее на расстоянии z. Очевидно, v есть функция от z ($z > 0$).

Чтобы узнать, чему равна функция v, попытаемся найти ее дифференциал dv. Приращение Δv функции v в точке z равно объему, ограниченному параболоидом и двумя плоскостями, параллельными плоскости Oxy и отстоящими от нее на расстояниях z и $z + \Delta z$.

Легко видеть, что величина Δv больше объема кругового цилиндра радиуса $\sqrt{z}$ и высоты Δz, но меньше объема кругового цилиндра радиуса $\sqrt{z + \Delta z}$ и высоты Δz. Таким образом,

$$\pi z \Delta z < \Delta v < \pi (z + \Delta z) \Delta z$$

и, следовательно,

$$\Delta v = \pi (z + \theta \Delta z) \Delta z = \pi z \Delta z + \pi \theta \Delta z^2,$$

где θ — некоторое число, зависящее от Δz и удовлетворяющее неравенству $0<\theta<1$.

Таким образом, нам удалось приращение Δv представить в виде суммы, первое слагаемое которой пропорционально Δz, а второе слагаемое есть бесконечно малая высшего порядка по отношению к Δz (при $\Delta z \to 0$). Отсюда следует, что первое слагаемое есть дифференциал функции v

$$dv = \pi z \, \Delta z,$$

или

$$dv = \pi z \, dz,$$

так как для независимой переменной z имеет место равенство $\Delta z = dz$.

Полученное равенство связывает между собой дифференциалы dv и dz (переменных v и z) и потому называется дифференциальным уравнением.

Если принять во внимание, что

$$\frac{dv}{dz} = v',$$

где v' есть производная от v по переменной z, то наше дифференциальное уравнение можно записать еще в виде

$$v' = \pi z.$$

Решение этого простейшего дифференциального уравнения сводится к отысканию функции от z, производная от которой равна πz. В общем виде подобной задаче посвящаются §§ 10 и 11, а сейчас мы предоставляем читателю проверить, что решением нашей задачи является функция $v = \frac{\pi z^2}{2} + C$, где в качестве C можно взять любое постоянное число[1]. В данном случае функция, выражающая объем рассматриваемого нами тела, очевидно, равна нулю при $z = 0$ (см. рис. 22), отсюда $C = 0$. Таким образом, наша функция определяется равенством $v = \frac{\pi z^2}{2}$.

Теорема о среднем и примеры ее применения. Дифференциал выражает приближенное значение приращения функции через приращение независимой переменной и производную в начальной точке. Если речь идет о приращении на участке от $x = a$ до $x = b$, то

$$f(b) - f(a) \approx f'(a)(b - a).$$

Можно получить точное равенство такого типа, если заменить справа производную $f'(a)$ в начальной точке на производную в некоторой средней, удачно подобранной точке промежутка (a, b). Точнее: *если $y = f(x)$ есть функция, дифференцируемая на отрезке $a \leqslant x \leqslant b$, то строго внутри его существует такая точка ξ, что имеет место точное равенство*

$$f(b) - f(a) = f'(\xi)(b - a). \tag{22}$$

[1] Это дает все решения (см. сноску на стр. 139).

Геометрический смысл этой «теоремы о среднем», известной под названием формулы Лагранжа или формулы конечных приращений, чрезвычайно прост. Пусть на графике функции $f(x)$ точки A и B, отвечающие значениям $x = a$ и $x = b$, соединены хордой AB (рис. 23). Будем переносить прямую AB, сдвигая ее параллельно самой себе вверх или вниз.

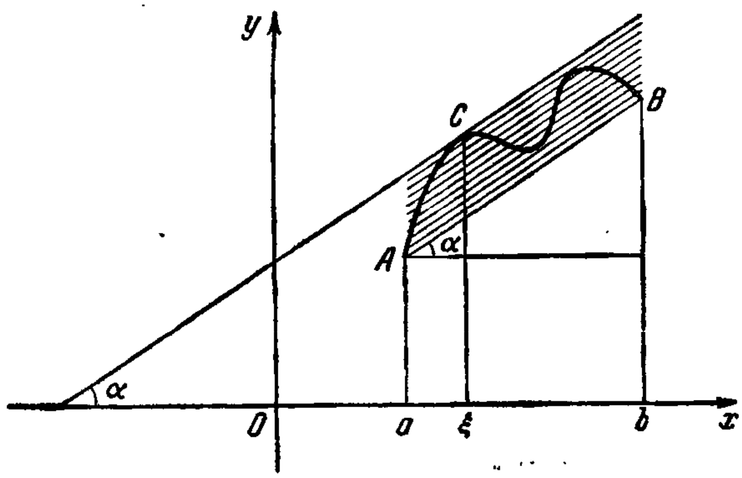


Рис. 23.

Тогда в тот момент, когда наша прямая будет пересекать график в последний раз, она будет касательной к нему в некоторой точке C. В этой точке (пусть ей соответствует $x = \xi$) касательная имеет тот же угол наклона α, что и хорда AB. Но у хорды

$$\operatorname{tg}\alpha = \frac{f(b)-f(a)}{b-a}.$$

С другой стороны, в точке C

$$\operatorname{tg}\alpha = f'(\xi).$$

Равенство

$$\frac{f(b)-f(a)}{b-a} = f'(\xi)$$

и выражает как раз теорему о среднем[1].

Формула (22) своеобразна тем, что в ней фигурирует неизвестная нам точка ξ, о которой мы знаем всякий раз только то, что она лежит «где-то в промежутке (a, b)». Несмотря на такую неопределенность, эта формула имеет большое теоретическое значение, являясь средством доказательства многих теорем анализа. Непосредственное практическое значение ее тоже велико, так как она дает возможность оценить приращение функции, когда известны пределы колебания ее производной. Например,

$$|\sin b - \sin a| = |\cos \xi|\,(b-a) \leqslant b-a.$$

Здесь a, b и ξ — углы, выраженные в радианах; ξ — некоторое значение между a и b; оно неизвестно, однако мы знаем, что $|\cos \xi| \leqslant 1$.

[1] Конечно, высказанные соображения лишь указывают на геометрический смысл теоремы, но не являются ее строгим доказательством.

Из формулы (22) очевидно, что функция, у которой производная все время равна нулю, должна быть постоянной, она ни на каком участке не может получить отличного от нуля приращения. Читатель аналогичным путем легко теперь докажет, что функция, производная которой все время положительна, обязательно возрастает, а при отрицательной производной убывает. Укажем без доказательства на одно из обобщений теоремы о среднем

Для любых дифференцируемых в $[a, b]$ функций $\varphi(x)$ и $\psi(x)$ [если только $\psi'(x) \neq 0$ в (a, b)] справедливо равенство[1]

$$\frac{\varphi(b)-\varphi(a)}{\psi(b)-\psi(a)} = \frac{\varphi'(\xi)}{\psi'(\xi)}, \tag{23}$$

где ξ — некоторая точка из промежутка (a, b)[2].

Из высказанного утверждения можно получить общий способ вычисления пределов вида

$$\lim_{x\to 0} \frac{\varphi(x)}{\psi(x)}, \tag{24}$$

если $\varphi(0) = \psi(0) = 0$. Пользуясь формулой (23), замечаем, что

$$\frac{\varphi(x)}{\psi(x)} = \frac{\varphi(x)-\varphi(0)}{\psi(x)-\psi(0)} = \frac{\varphi'(\xi)}{\psi'(\xi)},$$

где ξ находится между 0 и x, а потому $\xi \to 0$ вместе с x. Это позволяет вместо предела (24) искать $\lim\limits_{x\to 0} \frac{\varphi'(x)}{\psi'(x)}$, что в массе случаев чрезвычайно облегчает отыскание пределов[3].

Пример. Найти $\lim\limits_{x\to 0} \frac{x-\sin x}{x^3}$. Трижды применяя указанное правило, находим последовательно

$$\lim_{x\to 0} \frac{x-\sin x}{x^3} = \lim_{x\to 0} \frac{1-\cos x}{3x^2} = \lim_{x\to 0} \frac{\sin x}{6x} = \lim_{x\to 0} \frac{\cos x}{6} = \frac{1}{6}.$$

§ 9. ФОРМУЛА ТЕЙЛОРА

Функция

$$p(x) = a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n,$$

где коэффициенты a_k — постоянные числа, называется многочленом степени n. В частности, функция $y = ax + b$ есть многочлен 1-й степени,

[1] Формулу (23) можно получить простым применением теоремы о среднем к функции

$$f(x) = \varphi(x) - \frac{\varphi(b)-\varphi(a)}{\psi(b)-\psi(a)}\,\psi(x).$$

[2] Символами $[a, b]$ и (a, b) мы обозначаем, соответственно, множества значений x, удовлетворяющих неравенствам $a \leqslant x \leqslant b$ и $a < x < b$.

[3] То же правило верно и при отыскании предела дробного выражения, в котором числитель и знаменатель оба стремятся к бесконечности. Этот прием, весьма полезный для отыскания таких пределов (или, как еще говорят, для раскрытия неопределенностей), будет применяться, например, в § 3 главы XII (том 2).

а $y = ax^2 + bx + c$ — многочлен 2-й степени. Многочлены можно считать простейшими функциями. Для их вычисления по данному значению x достаточно действий сложения, вычитания и умножения, даже деления не требуется. Многочлены непрерывны при всяком x и имеют производные любого порядка. Кстати, производная от многочлена снова есть многочлен, на единицу меньшей степени, а производные порядка $n+1$ и выше от многочлена степени n равны нулю.

Если к многочленам присоединить функции вида

$$y = \frac{a_0 + a_1x + \cdots + a_nx^n}{b_0 + b_1x + \cdots + b_mx^m},$$

для вычисления которых уже требуется деление, и еще функции $\sqrt{x}$ и $\sqrt[3]{x}$, и, наконец, арифметические комбинации из таких функций, — вот, собственно говоря, и все функции, которые мы умеем вычислять с помощью методов, усвоенных из курса средней школы.

Уже на школьной скамье мы получили представление о ряде других функций, таких, как

$$\sqrt[5]{x},\ \lg x,\ \sin x,\ \operatorname{arctg} x,\ \ldots$$

Мы узнали важнейшие свойства этих функций, однако элементарная математика не дает ответа на вопрос: как вычислять их. Какие, например, действия надо совершить над x, чтобы получить $\lg x$ или $\sin x$. Ответ на этот вопрос дают методы, разработанные в анализе. На одном из таких методов мы остановимся здесь подробнее.

Формула Тейлора. Пусть на некотором промежутке, внутри которого лежит точка a, задана функция $f(x)$, имеющая производные всех порядков. Многочлен 1-й степени

$$p_1(x) = f(a) + f'(a)(x-a)$$

совпадает в точке $x=a$ с $f(x)$ и имеет в этой точке, как нетрудно проверить, ту же производную, что и $f(x)$. Его график является прямой, касательной к графику $f(x)$ в точке a. Можно подобрать многочлен 2-й степени, именно

$$p_2(x) = f(a) + f'(a)(x-a) + \frac{f''(a)}{2}(x-a)^2,$$

который в точке $x=a$ будет иметь с $f(x)$ общее значение и одинаковые первую и вторую производные. Его график будет вблизи точки a еще теснее прилегать к графику функции $f(x)$. Естественно ожидать, что если мы построим многочлен, имеющий при $x=a$ первые n производных, одинаковых с n производными $f(x)$ в той же точке, то этот многочлен при x, близких к a, будет лучше приближать $f(x)$. Так получается приближенное равенство, выражающее формулу Тейлора

$$f(x) \approx f(a) + f'(a)(x-a) + \frac{f''(a)}{2!}(x-a)^2 + \cdots + \frac{f^{(n)}(a)}{n!}(x-a)^n. \quad (25)$$

Правая часть этой формулы есть многочлен степени n от $(x - a)$. При каждом x его можно фактически вычислить, если известны $f(a)$, $f'(a)$, ..., $f^{(n)}(a)$.

Для функций, имеющих $n + 1$ производную, правая часть этой формулы, как это легко доказать, отличается от левой на малую величину, стремящуюся к нулю быстрее, чем $(x - a)^n$. Более того, это — единственный возможный многочлен степени n, отличающийся от $f(x)$ при x, близких к a, на величину, которая стремится при $x \to a$ к нулю быстрее, чем $(x - a)^n$. В случае же, когда $f(x)$ есть алгебраический многочлен степени n, приближенное равенство (25) обращается в точное.

Наконец, и это чрезвычайно важно, удается простым образом выразить, насколько именно отличается правая часть формулы (25) от $f(x)$. А именно, для того чтобы равенство (25) стало точным, надо добавить справа еще некоторый так называемый «остаточный член» формулы

$$f(x) = f(a) + f'(a)(x - a) + \cdots + \frac{f^{(n)}(a)}{n!}(x - a)^n + \frac{f^{(n+1)}(\xi)}{(n+1)!}(x - a)^{n+1}. \qquad (26)$$

Последнее, дополнительное, слагаемое[1]

$$R_{n+1}(x) = \frac{f^{(n+1)}(\xi)}{(n+1)!}(x - a)^{n+1}$$

имеет ту особенность, что стоящая в нем производная должна вычисляться всякий раз не в самой точке a, а в специально подобранной, заранее неизвестной нам точке ξ, заведомо лежащей, однако, где-то в промежутке между a и x.

Доказательство равенства (26) довольно громоздко, но весьма несложно по существу. Приведем несколько искусственный, но зато краткий вариант доказательства.

Чтобы установить, насколько отличается в приближенном равенстве (25) левая часть от правой, рассмотрим отношение разности левой и правой части равенства (25) к величине $-(x - a)^{n+1}$

$$\frac{f(x) - \left[f(a) + f'(a)(x - a) + \cdots + \dfrac{f^{(n)}(a)}{n!}(x - a)^n\right]}{-(x - a)^{n+1}}. \qquad (27)$$

Введем еще в рассмотрение функцию

$$\varphi(u) = f(u) + f'(u)(x - u) + \cdots + \frac{f^{(n)}(u)}{n!}(x - u)^n$$

от переменной u, считая, что x есть фиксированная (постоянная) величина. Тогда числитель (27) будет не чем иным, как приращением этой функции при переходе от $u = a$ к $u = x$, а знаменатель будет прираще-

[1] Это лишь одна из возможных форм выражения остаточного члена $R_{n+1}(x)$.

нием на том же участке функции

$$\psi(u)=(x-u)^{n+1}.$$

Остается воспользоваться известным из предыдущего параграфа обобщением теоремы о среднем

$$\frac{\varphi(x)-\varphi(a)}{\psi(x)-\psi(a)}=\frac{\varphi'(\xi)}{\psi'(\xi)}.$$

Выполняя дифференцирование по u функций $\varphi(u)$ и $\psi(u)$ (при этом надо помнить, что x постоянно: мы его зафиксировали), убедимся, что

$$\frac{\varphi'(\xi)}{\psi'(\xi)}=-\frac{f^{(n+1)}(\xi)}{(n+1)!}.$$

Равенство последнего выражения исходной величине (27) и представляет как раз формулу Тейлора в виде (26).

В последнем виде (26) формула Тейлора дает не только средство приближенно вычислять $f(x)$, но позволяет также оценивать допускаемую при этом погрешность.

Обратимся к простому примеру

$$y=\sin x.$$

Значения функции $\sin x$ и ее производных любого порядка при $x=0$ мы знаем. Воспользуемся этим и напишем формулу Тейлора для $\sin x$, полагая $a=0$ и ограничиваясь случаем $n=4$. Последовательно находим

$$\begin{array}{lll}
f(x)=\sin x, & f'(x)=\cos x, & f''(x)=-\sin x,\\
f'''(x)=-\cos x, & f^{\mathrm{IV}}(x)=\sin x, & f^{\mathrm{V}}(x)=\cos x;\\
f(0)=0, & f'(0)=1, & f''(0)=0,\\
f'''(0)=-1. & f^{\mathrm{IV}}(0)=0, & f^{\mathrm{V}}(\xi)=\cos\xi.
\end{array}$$

Поэтому

$$\sin x=x-\frac{x^3}{6}+R_5,\quad \text{где}\quad R_5=\frac{x^5}{120}\cos\xi.$$

Хотя точное значение R_5 нам неизвестно, но его легко оценить, учитывая, что $|\cos\xi|\leqslant 1$. Если ограничиться значениями x от 0 до $\frac{\pi}{4}$, то для таких x

$$|R_5|=\left|\frac{x^5}{120}\cos\xi\right|\leqslant\frac{1}{120}\left(\frac{\pi}{4}\right)^5<\frac{1}{400}.$$

Следовательно, на отрезке $\left[0,\frac{\pi}{4}\right]$ функцию $\sin x$ с точностью до $\frac{1}{400}$ можно считать равной многочлену 3-й степени

$$\sin x=x-\frac{1}{6}x^3.$$

Если в разложении $\sin x$ по формуле Тейлора взять больше членов, то получим многочлен более высокой степени, приближающий $\sin x$ еще точнее.

Подобными методами вычисляются тригонометрические и многие другие таблицы.

Законы природы, как правило, с хорошим приближением выражаются функциями, дифференцируемыми любое число раз, которые, в свою очередь, могут приближенно изображаться многочленами; выбор степени многочлена определяется необходимой точностью.

Ряд Тейлора. Если брать в формуле (25) все большее и большее число членов, то отклонение правой части от $f(x)$, выражаемое остаточным членом $R_{n+1}(x)$, может оказаться стремящимся к нулю. Конечно, это будет далеко не всегда: не для всякой функции и не для всякого x. Однако существует широкий класс функций (называемых *аналитическими*), для которых остаточный член $R_{n+1}(x)$ действительно стремится к нулю при $n \to \infty$, по крайней мере для значений x, заполняющих некоторый промежуток, окружающий точку a. Именно для таких функций формула Тейлора позволяет вычислять $f(x)$ с любой степенью точности. Остановимся подробнее на этих функциях.

Если $R_{n+1}(x) \to 0$ при $n \to \infty$, то из (26) следует

$$f(x) = \lim_{n \to \infty} \left[f(a) + f'(a)(x-a) + \cdots + \frac{f^{(n)}(a)}{n!}(x-a)^n \right].$$

В этом случае говорят, что $f(x)$ разлагается в сходящийся бесконечный ряд

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

расположенный по возрастающим степеням $(x-a)$ и называемый *рядом Тейлора*, причем $f(x)$ называют суммой этого ряда. Приведем несколько примеров разложений в ряд Тейлора хорошо известных нам функций (в этих примерах $a = 0$):

1) $(1+x)^n = 1 + nx + \frac{n(n-1)}{2!}x^2 + \frac{n(n-1)(n-2)}{3!}x^3 + \ldots$

(верно при $|x| < 1$ и любом действительном n).

2) $\sin x = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \ldots$ (верно при всех x).

3) $\cos x = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \frac{x^6}{6!} + \ldots$ (верно при всех x).

4) $e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \ldots$ (верно при всех x).

5) $\operatorname{arctg} x = x - \frac{x^3}{3} + \frac{x^5}{5} - \ldots$ (верно при $|x| \leqslant 1$).

Первый из этих примеров — известный бином Ньютона, полученный Ньютоном для всех n, но в его время полностью обоснованный только для целых n. Этот пример послужил образцом к созданию общей формулы Тейлора. Последние два примера при $x = 1$ позволяют вычислить с любой степенью точности числа e и π.

Практическое значение формулы Тейлора, открывающей путь большинству вычислений в приложениях анализа, весьма велико.

Функции, разлагающиеся в ряд Тейлора, с большой точностью выражают многие закономерности в природе: физические и химические процессы, движения тел и т. д. Их теория получает наиболее ясный и законченный характер, если рассматривать их как функции комплексного переменного. Теории таких функций будет уделено соответствующее место в главе IX (том 2).

Сама идея приближенного выражения функции многочленом и задача представления функции в виде суммы бесконечного числа более простых слагаемых получили далеко идущее развитие в анализе, образовав самостоятельный его раздел — теорию приближения функций (см. главу XII, том 2).

§ 10. ИНТЕГРАЛ

Из главы 1 и из § 1 настоящей главы читатель уже знает, что понятие интеграла и вообще интегральное исчисление исторически возникли из необходимости решения конкретных задач, характерным примером которых является задача о нахождении площади криволинейной фигуры. Настоящий параграф посвящен этим вопросам. Из него мы узнаем также, в чем заключается связь между задачами дифференциального и интегрального исчисления, которая полностью была выяснена только в XVIII в.

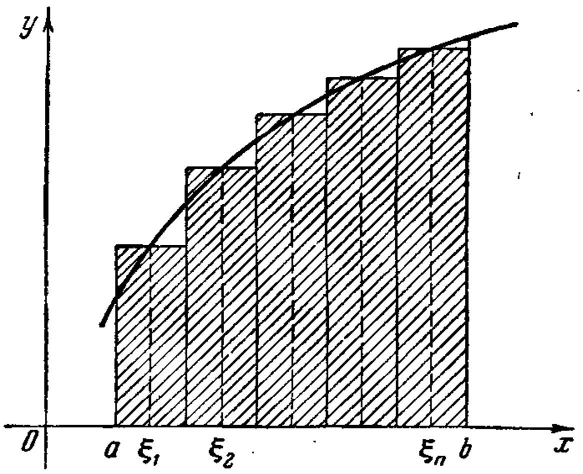


Рис. 24.

Площадь. Пусть проходящая над осью x-ов линия служит графиком функции $y = f(x)$. Попытаемся найти площадь S участка, ограниченного линией $y = f(x)$, осью x-ов и прямыми, проведенными через точки $x = a$ и $x = b$ параллельно оси y-ов.

Для решения этой задачи поступим следующим образом. Разобьем отрезок $[a, b]$ на n частей (не обязательно одинаковых). Обозначим длину первого участка Δx_1, второго Δx_2 и т. д., последнего Δx_n. В каждом участке выберем по точке $\xi_1, \xi_2, \ldots, \xi_n$, после чего составим сумму

$$S_n = f(\xi_1)\Delta x_1 + f(\xi_2)\Delta x_2 + \ldots + f(\xi_n)\Delta x_n. \tag{28}$$

Величина S_n, очевидно, равна сумме площадей заштрихованных на рис. 24 прямоугольников.

Чем мельче сделать разбиение участка $[a, b]$, тем ближе будет S_n к площади S. Если производить ряд таких построений, разбивая промежуток $[a, b]$ на все более и более мелкие участки, то суммы S_n будут стремиться к S.

Возможность деления $[a, b]$ на неравные участки заставляет нас уточнить, что следует понимать под «все более мелкими» разбиениями. Мы предполагаем, что не только n неограниченно возрастает, но даже наибольшая из длин Δx_i в n-м разбиении стремится к нулю. Итак

$$S = \lim_{\max \Delta x_i \to 0} [f(\xi_1)\Delta x_1 + f(\xi_2)\Delta x_2 + \ldots + f(\xi_n)\Delta x_n] =$$

$$= \lim_{\max \Delta x_i \to 0} \sum_{i=1}^{n} f(\xi_i)\Delta x_i. \qquad (29)$$

Вычисление площади свелось к отысканию предела (29).

Отметим, что когда мы ставили нашу задачу, то имели только опытное представление о площади нашей криволинейной фигуры, но точного определения не имели. В результате проведенных рассуждений мы получили точное определение понятия площади: это есть предел (29). Теперь мы не только имеем наглядное представление о площади, но располагаем и ее математическим определением, позволяющим вычислять площадь (сравните с замечаниями на стр. 99 о длине окружности и скорости).

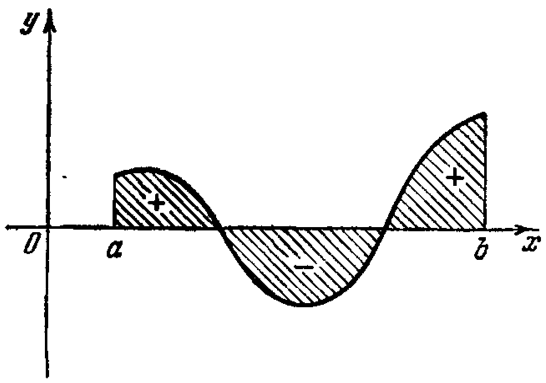


Рис. 25.

Мы предполагали, что $f(x) \geqslant 0$. Если $f(x)$ меняет знак, как на рис. 25, то предел (29) дает алгебраическую сумму площадей участков, лежащих между кривой $y = f(x)$ и осью x-ов, причем площади участков, лежащих над осью x-ов, считаются со знаком плюс, а участков, лежащих под осью,— со знаком минус.

Определенный интеграл. К необходимости вычислять предел (29) приводит также много других задач. Пусть, например, точка движется по прямой с переменной скоростью $v = f(t)$. Как определить путь s, пройденный ею за промежуток времени от $t = a$ до $t = b$?

Будем предполагать, что функция $f(t)$ непрерывна, т. е., что в малые промежутки времени скорость мало изменяется. Разделим $[a, b]$ на n частей, длительности $\Delta t_1, \Delta t_2, \ldots, \Delta t_n$. Чтобы приближенно вычислить путь, пройденный за каждый промежуток Δt_i, будем считать, что скорость в этот период времени была постоянной, такой, как в некоторый произвольный момент ξ_i из этого промежутка. Весь пройденный путь приближенно выразится тогда суммой

$$s_n = \sum_{i=1}^{n} f(\xi_i)\Delta t_i,$$

а точное значение пути s, пройденного за время от a до b, мы найдем, как предел таких сумм при все более мелких разбиениях,— это будет предел вида (29)

$$s = \lim_{\max \Delta t_i \to 0} \sum_{i=1}^{n} f(\xi_i)\, \Delta t_i.$$

Можно привести много конкретных задач, решение которых сводится к вычислению подобного предела. Мы еще встретимся с ними, но уже эти примеры достаточно выясняют его важность. Предел (29) называют *определенным интегралом* функции $f(x)$, взятым по отрезку $[a, b]$, и обозначают через

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

Выражение $f(x)\, dx$ называется подинтегральным выражением, a и b — пределами интегрирования; a — нижний предел, b — верхний предел.

Связь дифференциального и интегрального исчисления. Примером прямого вычисления определенного интеграла может служить пример 2 § 1. Теперь можно сказать, что рассмотренная там задача сводится к вычислению определенного интеграла

$$\int_0^h ax\, dx.$$

Другой пример был рассмотрен в § 3, где была решена задача о нахождении площади, ограниченной параболой $y = x^2$. Здесь вопрос сводится к вычислению интеграла

$$\int_0^1 x^2\, dx.$$

Оба эти интеграла нам удалось вычислить благодаря тому, что мы знали простые формулы для суммы первых n натуральных чисел и суммы квадратов этих чисел. Однако далеко не для всякой функции $f(x)$ удается суммировать (выражать простыми формулами) суммы (28), где точки ξ_i и приращения Δx_i заданы тем или иным законом. Больше того, в тех случаях, когда такое суммирование возможно, оно осуществляется не при помощи одного общего метода, а при помощи различных методов весьма частного вида, присущих чуть ли не каждой задаче в отдельности.

Таким образом, возникает задача указать общий метод вычисления определенных интегралов. Исторически эта задача долгое время стояла перед математиками в виде конкретной проблемы отыскания общего метода нахождения площадей криволинейных фигур, объемов тел, ограниченных кривой поверхностью, и т. д.

Мы уже отмечали, что Архимед умел вычислять площадь сегмента и некоторых других фигур. В дальнейшем число отдельных задач подобного рода на вычисление площадей, объемов, центров тяжести тел и т. д., которые научились решать, постепенно увеличивалось. Однако процесс создания общего метода решения этих задач происходил сначала весьма медленно. Общий метод мог быть создан только после того, как был накоплен достаточно большой теоретический и вычислительный материал, который, в свою очередь, создавался в теснейшей связи с потребностями практики. Процесс накопления и обобщения сильно ускорился и перешел в энергичное развитие только в конце средних веков, что являлось непосредственным следствием бурного развития производительных сил в Европе в ту эпоху, знаменующую ломку прежних (феодальных) производственных отношений и создание новых (капиталистических).

Накопление фактов, связанных с задачей вычисления определенных интегралов, происходило параллельно с соответствующими исследованиями, связанными с задачей о нахождении производной от функции. Читатель уже знает из § 1, что эта огромная подготовительная работа получила завершение в XVII в. в работах Ньютона и Лейбница. В этом смысле говорят, что Ньютон и Лейбниц являются создателями дифференциального и интегрального исчислений.

Одна из основных заслуг Ньютона и Лейбница заключается в том, что в их работах окончательно была выяснена та глубокая связь, которая существует между дифференциальным и интегральным исчислениями. Эта связь, в частности, дает общий метод вычисления определенных интегралов для весьма большого класса функций.

Чтобы разъяснить эту связь, мы обратимся к примеру из механики.

Допустим, что некоторая материальная точка движется по прямой со скоростью $v = f(t)$, где t — время. Мы уже знаем, что путь σ, пройденный нашей точкой за промежуток времени между $t = t_1$ и $t = t_2$, равен определенному интегралу

$$\sigma = \int_{t_1}^{t_2} f(t)\,dt.$$

Кроме того, допустим, что нам известен закон движения точки, т. е. известна функция $s = F(t)$, выражающая зависимость от времени t пути s, исчисляемого от некоторой начальной точки A, выбранной на прямой. Пройденный за промежуток времени $[t_1, t_2]$ путь σ, очевидно, равен разности

$$\sigma = F(t_2) - F(t_1).$$

Таким образом из физических соображений мы пришли к равенству

$$\int_{t_1}^{t_2} f(t)\,dt = F(t_2) - F(t_1),$$

которое выражает связь между законом движения нашей точки и ее скоростью.

С математической точки зрения функция $F(t)$, как это мы знаем из § 5, может быть определена как такая функция, производная от которой для всех рассматриваемых t равна $f(t)$, т. е.

$$F'(t) = f(t).$$

Такая функция называется *первообразной* функцией по отношению к $f(t)$.

Надо иметь в виду, что если функция $f(t)$ имеет хоть одну первообразную, то вместе с ней она имеет и бесчисленное множество первообразных, потому что если $F(t)$ есть первообразная для $f(t)$, то $F(t) + C$, где C — произвольная постоянная, есть также первообразная для $f(t)$. Однако этим исчерпывается вся совокупность первообразных функций для $f(t)$, так как если $F_1(t)$ и $F_2(t)$ являются первообразными для одной и той же функции $f(t)$, то их разность $\varphi(t) = F_1(t) - F_2(t)$ имеет всюду на рассматриваемом отрезке изменения t производную $\varphi'(t)$, равную нулю и, следовательно, есть постоянная [1].

С физической точки зрения различные значения постоянной C определяют законы движения, отличающиеся только тем, что они соответствуют всевозможным различным начальным точкам O отсчета пути.

Сказанное приводит к выводу, что при весьма общих условиях, накладываемых на функцию $f(x)$, заданную на отрезке $[a, b]$, во всяком случае при таких условиях, когда функцию $f(x)$ можно рассматривать как скорость движения некоторой точки в момент времени x, справедливо равенство [2]

$$\int_a^b f(x)\,dx = F(b) - F(a), \tag{30}$$

где $F(x)$ — любая первообразная функция для $f(x)$.

Это равенство и есть знаменитая *формула Ньютона и Лейбница*, сводящая вопрос о вычислении определенного интеграла от функции к отысканию ее первообразной и, таким образом, связывающая в себе дифференциальное и интегральное исчисления.

Многие частные задачи, которые были предметом исследования крупнейших математиков, автоматически решаются при помощи этой формулы, выражающей, что определенный интеграл от функции $f(x)$ на отрезке $[a, b]$ равен разности значений какой-либо первообразной для этой

---

[1] По теореме о среднем

$$\varphi(t) - \varphi(t_0) = \varphi'(\nu)(t - t_0) = 0,$$

где ν находится между t и t_0. Отсюда $\varphi(t) = \varphi(t_0) = \text{const}$ для всех t.

[2] Можно математически, не прибегая к механическим примерам, доказать, что если функция $f(x)$ непрерывна (и даже разрывна, но суммируема по Лебегу, см. главу XV, том 3) на отрезке $[a, b]$, то для нее существует первообразная $F(x)$ и выполняется равенство (30).

функции, соответствующих правому и левому концам отрезка [1]. Разность (30) еще принято записывать так:

$$F(x)\Big|_a^b = F(b) - F(a).$$

Пример 1. Равенство

$$\left(\frac{x^3}{3}\right)' = x^2$$

показывает, что функция $\frac{x^3}{3}$ есть первообразная для функции x^2. Отсюда, на основании формулы Ньютона и Лейбница,

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

Пример 2. Пусть c и c' — два заряда, находящиеся на прямой на расстоянии r друг от друга. Сила взаимодействия F между ними направлена вдоль этой прямой и равна

$$F = \frac{a}{r^2}$$

($a = kcc'$, где k — постоянная). Работу W этой силы, когда заряд c неподвижен, а заряд c' передвигается по отрезку $[R_1, R_2]$, можно подсчитать, разбивая отрезок $[R_1, R_2]$ на части Δr_i. На каждой из них приближенно считаем силу постоянной, тогда работа на таком участке равна $\frac{a}{r_i^2}\Delta r_i$. Делая частицы разбиения все более мелкими, убеждаемся, что работа W равна интегралу

$$W = \lim_{n\to\infty}\sum_{i=1}^{n}\frac{a}{r_i^2}\Delta r_i = \int_{R_1}^{R_2}\frac{a}{r^2}\,dr.$$

Этот интеграл мы сразу находим, принимая во внимание, что

$$\frac{a}{r^2} = \left(-\frac{a}{r}\right)',$$

отсюда

$$W = -\frac{a}{r}\Big|_{R_1}^{R_2} = a\left(\frac{1}{R_1} - \frac{1}{R_2}\right).$$

В частности, работа, выполненная силой F при передвижении заряда c', находящегося сначала на расстоянии R_1 от заряда c, на бесконечность, равна

$$W = \lim_{R_2\to\infty} a\left(\frac{1}{R_1} - \frac{1}{R_2}\right) = \frac{a}{R_1}.$$

Уже из соображений, при помощи которых мы пришли к формуле Ньютона и Лейбница, видно, что эта формула математически выражает определенную глубокую связь, имеющую место в объективной действи-

[1] Эта формула получила различные обобщения (см., например, § 13, формула Остроградского).

тельности. Формула Ньютона и Лейбница является прекрасным и притом весьма важным примером того, как математика отражает в себе объективные закономерности.

Нужно сказать, что Ньютон в своих математических исследованиях стоял на физической точке зрения. Его работы по созданию основ дифференциального и интегрального исчисления не отделимы от его работ по созданию основ механики.

Сами понятия математического анализа такие, как производная, интеграл, в представлениях Ньютона и его современников еще окончательно «не оторвались» от их физических и геометрических прообразов (скорость, площадь). Они в сущности носили наполовину математический и наполовину физический характер. Дело в том, что существовавшие тогда определения этих понятий с математической точки зрения были еще неудовлетворительными. Поэтому правильное оперирование ими в сколько-нибудь сложных случаях требовало от исследователя умения не отрываться от конкретной стороны вопроса даже на промежуточных стадиях рассуждений.

С этой точки зрения характер творчества Ньютона и Лейбница был различным [1]. Ньютоном на всех стадиях его исследований всегда руководила физическая точка зрения. Исследования же Лейбница не имеют такой тесной непосредственной связи с физикой, что при отсутствии четких математических определений приводило его иногда на отдельных стадиях исследований к ошибочным заключениям. С другой стороны, для творчества Лейбница было особенно характерно стремление к общности, стремление отыскивать возможно более общие методы решения задач математического анализа.

Важнейшей заслугой Лейбница является создание отражающей существо дела математической символики. Обозначения таких, например, основных понятий математического анализа, как дифференциала dx, второго дифференциала d^2x, интеграла $\int y\,dx$, производной $\frac{d}{dx}$ были предложены Лейбницем. Тот факт, что эти обозначения употребляются и поныне, свидетельствует о том, насколько они удачны.

Хорошо выбранная символика весьма способствует быстроте и легкости наших выкладок и рассуждений. Больше того, она подчас ограждает нас от ошибочных заключений. Лейбниц, который это хорошо сознавал, уделял в своем творчестве очень большое внимание выбору обозначений.

Эволюция понятий математического анализа (производной, интеграла и т. д.) происходила, конечно, и после Ньютона и Лейбница и происходит до наших дней. Однако важно отметить определенный этап в этой эволюции, который произошел в начале прошлого столетия и связан прежде всего с работами Коши.

[1] Открытия Ньютона и Лейбница происходили независимо друг от друга.

Коши дал четкое формальное определение понятия предела и на его основе — определения понятий непрерывности, производной, дифференциала, интеграла. Эти определения приведены в соответствующих местах настоящей главы. Ими мы широко пользуемся в современном анализе.

Важность этих достижений заключается в том, что оказалось возможным оперировать чисто формально не только в арифметике, алгебре, элементарной геометрии, но еще и в новой весьма широкой области математики, в математическом анализе, получая при этом правильные результаты.

По отношению к применению во всяком случае основной массы результатов математического анализа можно было теперь сказать: если исходные данные практически верны, то результаты математических рассуждений тоже практически верны; если мы убеждены в достаточной точности исходных данных, то справедливость полученных результатов нет необходимости проверять на практике, для этого достаточно убедиться только в правильности формальных рассуждений.

Сказанное требует, разумеется, следующей оговорки. В математических рассуждениях исходные данные, которые мы берем из практики, верны лишь с точностью до некоторых погрешностей. Это приводит к тому, что на каждом шагу наших математических рассуждений с практическими данными в свою очередь получаются результаты с определенными погрешностями, накапливающимися при этом по мере увеличения числа шагов [1].

Возвращаясь к определенным интегралам, остановимся на вопросе, имеющем принципиальное значение. Для каких функций $f(x)$, заданных на отрезке $[a, b]$, можно гарантировать существование определенного интеграла $\int_a^b f(x)\,dx$, т. е. числа, к которому стремится сумма $\sum_1^n f(\xi_i)\,\Delta x_i$, когда $\max \Delta x_i \to 0$? Имеется в виду, что это число должно быть одним и тем же при любом способе разбиения отрезка $[a, b]$ и любом способе выбора точек ξ_i.

Функции, для которых определенный интеграл, т. е. предел (29), действительно существует, называются интегрируемыми на отрезке $[a, b]$. Соответствующие исследования, относящиеся к прошлому веку, показали, что все непрерывные функции интегрируемы.

Имеются и разрывные интегрируемые функции. К их числу, например, относятся возрастающие и убывающие на отрезке $[a, b]$ ограниченные функции.

---

[1] Например, формально из $a = b$ и $b = c$ следует $a = c$. На практике же эта связь выглядит так: из того, что $a = b$ с точностью до ε и $b = c$ с точностью до ε, следует, что $a = c$ с точностью до 2ε.

Функция, равная нулю в рациональных точках$[a, b]$ и единице в иррациональных, может служить примером неинтегрируемой функции, так как при любом разбиении отрезка интегральная сумма s_n будет равняться нулю или единице, в зависимости от того, будем ли мы выбирать в качестве ξ_i рациональные числа или иррациональные.

Заметим, что на вопрос, как фактически находить определенный интеграл, во многих случаях дает ответ формула Ньютона — Лейбница. Однако здесь возникает проблема нахождения по данной функции ее первообразной, т. е. функции, имеющей данную функцию своей производной. К этой проблеме мы и перейдем. Заметим кстáти, что отыскание первообразной имеет большое значение и в других вопросах математики, особенно при решении дифференциальных уравнений.

§ 11. НЕОПРЕДЕЛЕННЫЕ ИНТЕГРАЛЫ. ТЕХНИКА ИНТЕГРИРОВАНИЯ

В математике принято произвольную первообразную функцию для заданной функции $f(x)$ называть *неопределенным интегралом* от $f(x)$ и обозначать его в виде

$$\int f(x)\,dx.$$

Таким образом, если $F(x)$ есть некоторая вполне определенная первообразная для $f(x)$, то неопределенный интеграл от $f(x)$ равен

$$\int f(x)\,dx = F(x) + C, \tag{31}$$

где C — произвольная постоянная.

Отметим еще, что если функция $f(x)$ дана на отрезке $[a, b]$, $F(x)$—ее первообразная и x — точка отрезка $[a, b]$, то на основании формулы Ньютона — Лейбница можно написать

$$F(x) = F(a) + \int_a^x f(t)\,dt.$$

Таким образом интеграл, стоящий в правой части этого равенства, только на постоянную $F(a)$ отличается от функции $F(x)$ первообразной для $f(x)$. В таком случае этот интеграл, если его рассматривать как функцию верхнего предела x (при переменном x), есть некоторая вполне определенная первообразная для $f(x)$ и, следовательно, неопределенный интеграл от $f(x)$ можно записать еще в следующем виде:

$$\int f(x)\,dx = \int_a^x f(t)\,dt + C,$$

где C — произвольная постоянная.

Приведем основную таблицу неопределенных интегралов, составленную непосредственно из соответствующей таблицы производных (см. § 6):

$$\int x^a\,dx = \frac{x^{a+1}}{a+1} + C\ (a \neq -1), \qquad \int \cos x\,dx = \sin x + C,$$

$$\int \frac{dx}{x} = \ln|x| + C\,^1, \qquad \int \sec^2 x\,dx = \operatorname{tg} x + C,$$

$$\int a^x\,dx = \frac{a^x}{\ln a} + C, \qquad \int \frac{dx}{\sqrt{1-x^2}} = \arcsin x + C = \tag{32}$$

$$\int e^x\,dx = e^x + C, \qquad = -\arccos x + C_1\ \left(C_1 - C = \frac{\pi}{2}\right),$$

$$\int \sin x\,dx = -\cos x + C, \qquad \int \frac{dx}{1+x^2} = \operatorname{arctg} x + C.$$

Общие свойства неопределенных интегралов также выводятся на основании соответствующих свойств производных. Например, из правила дифференцирования суммы получаем формулу

$$\int [f(x) \pm \varphi(x)]\,dx = \int f(x)\,dx \pm \int \varphi(x)\,dx + C,$$

а из соответствующего правила, выражающего, что постоянный множитель k можно выносить за знак производной, получим

$$\int kf(x)\,dx = k\int f(x)\,dx + C.$$

Таким образом,

$$\int \left(3x^2 + 2x - \frac{3}{\sqrt{x}} + \frac{4}{x} - 1\right) dx = 3\,\frac{x^3}{3} + \frac{2x^2}{2} - 3\,\frac{x^{-1/2+1}}{-\frac{1}{2}+1} + 4\ln|x| - x + C.$$

Существует ряд методов вычисления неопределенных интегралов. Остановимся на одном из них, а именно на *методе подстановки* или замены переменной, который основан на справедливости следующего равенства:

$$\int f(x)\,dx = \int f[\varphi(t)]\,\varphi'(t)\,dt + C, \tag{33}$$

$x = \varphi(t)$ — дифференцируемая функция. Соотношение (33) надо понимать в том смысле, что если и функции

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

равной левой части равенства (33), положить $x = \varphi(t)$, то получим такую функцию $F[\varphi(t)]$, производная от которой по t равна выражению, стоящему под знаком интеграла в правой части равенства (33). Это непосредственно следует из теоремы о производной функции от функции.

Приведем несколько примеров на применение метода подстановки

$$\int e^{kx}\,dx = \int e^t\,\frac{1}{k}\,dt = \frac{1}{k}\int e^t\,dt = \frac{1}{k}\,e^t + C = \frac{e^{kx}}{k} + C$$

(подстановка $kx = t$, откуда $k\,dx = dt$).

[1] При $x > 0$ $(\ln|x|)' = (\ln x)' = \frac{1}{x}$,

при $x < 0$ $(\ln|x|)' = [\ln(-x)]' = \frac{1}{-x}(-1) = \frac{1}{x}$.

$$\int \frac{x\,dx}{\sqrt{a^2-x^2}} = -\int dt = -t + C = -\sqrt{a^2-x^2} + C$$

$\left(\text{подстановка} \quad t = \sqrt{a^2-x^2}, \text{ откуда } dt = -\frac{x\,dx}{\sqrt{a^2-x^2}}\right)$.

$$\int \sqrt{a^2-x^2}\,dx = \int \sqrt{a^2-a^2\sin^2 u}\; a\cos u\,du = a^2\int \cos^2 u\,du =$$

$$= a^2\int \frac{1+\cos 2u}{2}\,du = \frac{a^2}{2}\left(u + \frac{\sin 2u}{2}\right) + C =$$

$$= \frac{a^2}{2}(u + \sin u\cos u) + C = \frac{a^2}{2}\left(\arcsin\frac{x}{a} + \frac{x}{a^2}\sqrt{a^2-x^2}\right) + C$$

(подстановка $x = a\sin u$).

Как видно из примеров, метод замены переменных значительно расширяет класс тех элементарных функций, которые мы теперь можем проинтегрировать, т. е. получить для них первообразные, являющиеся снова элементарными функциями. Однако надо иметь в виду, что с вычислительной точки зрения с интегрированием обстоит дело, вообще говоря, гораздо хуже, чем с дифференцированием.

Из § 6 известно, что производная от любой элементарной функции есть снова элементарная функция, которую можно получить совершенно эффективно, воспользовавшись правилами дифференцирования. Но обратное утверждение, вообще говоря, неверно, так как существуют такие элементарные функции, неопределенные интегралы от которых не являются в свою очередь элементарными функциями. Например, такими функциями являются e^{-x^2}, $\frac{1}{\ln x}$, $\frac{\sin x}{x}$ и др. Для получения интегралов от них приходится пользоваться приближенными методами, а также вводить в обиход новые функции, не сводимые к элементарным. Мы не имеем возможности задерживаться на этом вопросе, заметим только, что уже в элементарной математике можно найти много примеров, когда прямая операция выполнима в некотором классе чисел, в то время как обратная ей операция в этом же классе не выполняется; так, квадрат любого положительного рационального числа есть снова рациональное число, но корень квадратный из рационального числа далеко не всегда является таким. Аналогично дифференцирование элементарных функций дает функцию вновь элементарную, а интегрирование может вывести нас из этого класса функций.

Некоторые интегралы, не вычисляемые в элементарных функциях, имеют большое значение в математике и ее приложениях. Таким, например, является интеграл

$$\int_0^x e^{-t^2}\,dt,$$

который играет очень важную роль в теории вероятностей (см. главу XI, том 2). Укажем еще на интегралы

$$\int_0^\varphi \frac{d\theta}{\sqrt{1-k^2\sin^2\theta}} \quad \text{и} \quad \int_0^\varphi \sqrt{1-k^2\sin^2\theta}\,d\theta \qquad (k^2<1),$$

носящие название *эллиптических интегралов* соответственно первого и второго рода. К вычислению их сводится очень большое число задач механики и физики (см. главу V (том 2), § 1, пример 3). Составлены подробные таблицы значений этих интегралов для различных значений аргументов x и φ, вычисленные приближенными методами, но с большой точностью.

Надо подчеркнуть, что доказательство самого факта, что та или иная элементарная функция не интегрируется в элементарных же функциях, в каждом отдельном случае представляет большие трудности. Эти вопросы, исследование которых сыграло важную роль в развитии анализа, занимали умы выдающихся математиков-аналитиков прошлого века. Фундаментальные результаты принадлежат здесь Чебышеву, который, в частности, полностью исследовал вопрос о возможности интегрирования в элементарных функциях интеграла вида

$$\int x^m (a + bx^s)^p \, dx,$$

где m, s и p — рациональные числа. До Чебышева были известны полученные еще Ньютоном три соотношения между показателями m, s и p, которые влекут за собой интегрируемость в элементарных функциях этого интеграла. П. Л. Чебышев показал, что во всех остальных случаях этот интеграл в элементарных функциях не выражается.

Приведем еще один метод интегрирования — метод интегрирования по частям. Он основан на известной читателю формуле

$$(uv)' = uv' + u'v\ ,$$

производной от произведения функций u и v. Ее можно записать еще так:

$$uv' = (uv)' - u'v.$$

Теперь проинтегрируем левую и правую части и примем во внимание, что

$$\int (uv)' dx = uv + C,$$

тогда окончательно получим равенство

$$\int uv' dx = uv - \int u'v \, dx,$$

которое и называется *формулой интегрирования по частям*. (Постоянную C мы не написали, так как можно считать, что она включена в один из входящих в равенства неопределенных интегралов.)

Приведем примеры применения этой формулы. Требуется вычислить $\int xe^x \, dx$. В нем будем считать $u = x$ и $v' = e^x$, тогда $u' = 1$, $v = e^x$, и, следовательно:

$$\int xe^x \, dx = xe^x - \int 1 \cdot e^x \, dx = xe^x - e^x + C.$$

В интеграле $\int \ln x \, dx$ удобно считать $u = \ln x$, $v' = 1$, тогда $u' = \frac{1}{x}$, $v = x$ и

$$\int \ln x \, dx = x \ln x - \int dx = x \ln x - x + C.$$

Еще характерный пример, где приходится интегрировать по частям два раза, а затем искомый интеграл находить из полученного уравнения

$$\int e^x \sin x\,dx = e^x \sin x - \int e^x \cos x\,dx = e^x \sin x - e^x \cos x - \int e^x \sin x\,dx,$$

откуда

$$\int e^x \sin x\,dx = \frac{e^x}{2}(\sin x - \cos x) + C.$$

Этим мы заканчиваем параграф, из которого читатель получил лишь поверхностное представление о теории интегрирования. На многих методах этой теории мы не остановились. В частности, мы не коснулись здесь очень интересной теории интегрирования рациональных дробей — теории, в которую внес важный вклад известный математик и механик прошлого века М. В. Остроградский.

§ 12. ФУНКЦИИ МНОГИХ ПЕРЕМЕННЫХ

До сих пор мы говорили о функциях одной переменной, но на практике часто приходится иметь дело также с функциями, зависящими от двух, трех и вообще многих переменных. Например, площадь прямоугольника есть функция

$$S = xy$$

его основания x и высоты y. Объем прямоугольного параллелепипеда есть функция

$$v = xyz$$

от трех его измерений. Расстояние между двумя точками A и B есть функция

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

от шести координат этих точек. Известная формула

$$pv = RT$$

выражает зависимость объема v определенного количества газа от его давления p и абсолютной температуры T.

Функции многих переменных, как и функции одной переменной, задаются обычно лишь для некоторой области значений самих переменных. Например, функция

$$u = \ln(1 - x^2 - y^2 - z^2) \tag{34}$$

задана только при значениях x, y, z, удовлетворяющих условию

$$x^2 + y^2 + z^2 < 1. \tag{35}$$

(При других x, y, z ее значения не являются действительными числами). Совокупность точек пространства, координаты которых удовлетворяют неравенству (35), заполняет, очевидно, шар единичного радиуса с центром в начале координат. Точки границы не причисляются к шару, с него как бы «содрана» его поверхность. Такой шар называют открытым. Функция (34) определена только для таких троек чисел (x, y, z), которые являются координатами точек этого открытого шара G. Коротко принято говорить, что функция (34) определена на шаре G.

Вот еще другой пример. Температура неравномерно нагретого тела V есть некоторая функция координат x, y, z точек этого тела. Эта функция задана не для всех троек чисел x, y, z, а только для таких троек, которые являются координатами точек нашего тела V.

Наконец, в качестве третьего примера рассмотрим функцию

$$u = \varphi(x) + \varphi(y) + \varphi(z),$$

где φ — функция одной переменной, определенная на отрезке $[0, 1]$. Очевидно функция u задана только для троек (x, y, z) чисел, являющихся координатами точек, заполняющих куб:

$$0 \leqslant x \leqslant 1, \quad 0 \leqslant y \leqslant 1, \quad 0 \leqslant z \leqslant 1.$$

Дадим определение функции трех переменных. Пусть задано множество E троек чисел (x, y, z) (точек пространства). Если каждой тройке (точке) E в силу некоторого закона соответствует определенное число u, то говорят, что u есть функция от x, y, z (от точки), определенная на множестве троек (точек) E, что записывается так:

$$u = F(x, y, z).$$

Вместо F можно писать другие буквы: f, φ, ψ.

На практике в качестве множества E фигурируют обычно множества точек, заполняющих некоторое геометрическое тело (область): шар, куб, кольцо и т. д., и тогда просто говорят, что функция определена на этом теле (области). Аналогично определяются функции двух, четырех и т. д. переменных.

Неявное задание функции. Отметим, что функции двух переменных при известных обстоятельствах могут служить хорошим средством задания функции одной переменной. Зададим функцию $F(x, y)$ от двух переменных и составим уравнение

$$F(x, y) = 0. \tag{36}$$

Оно, вообще говоря, определяет некоторое множество точек (x, y) плоскости (x, y), на котором наша функция равна нулю. Часто такие множества представляют собой некоторые кривые, которые можно рассматривать как графики одной или нескольких однозначных функций $y = \varphi(x)$ или $x = \psi(y)$ от одной переменной. В таком случае говорят, что эти однозначные функции определяются неявно при помощи уравнения (36). Например, уравнение

$$x^2 + y^2 - r^2 = 0$$

определяет неявным образом две функции от одной переменной

$$y = +\sqrt{r^2 - x^2} \text{ и } y = -\sqrt{r^2 - x^2}.$$

Впрочем, надо иметь в виду, что уравнение вида (36) может и не определять совсем никакой функции. Например, уравнение

$$x^2 + y^2 + 1 = 0,$$

очевидно, никакой действительной функции не задает, так как никакая пара действительных чисел этому уравнению не удовлетворяет.

Геометрическое изображение. Функции двух переменных можно весьма наглядно изображать в виде поверхностей с помощью пространственной системы координат. Так, функция

$$z = f(x, y) \tag{37}$$

изображается в трехмерной прямоугольной системе координат поверхностью, представляющей собою геометрическое место точек M, координаты x, y, z которых удовлетворяют уравнению (37) (рис. 26).

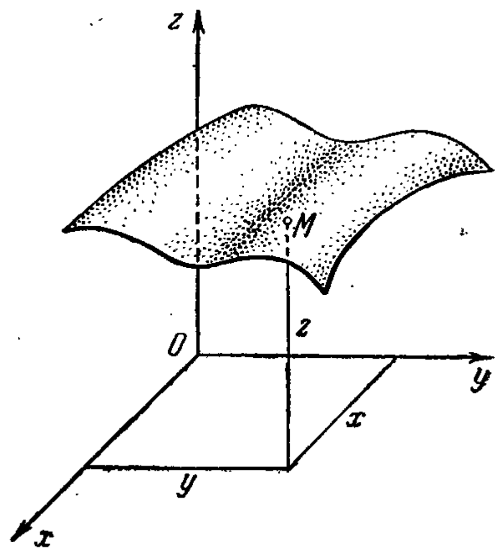


Рис. 26.

Существует еще другой, весьма удобный способ изображения функции (37), получивший широкое применение на практике. Задаваясь рядом значений $z_1, z_2, \ldots$, чертят в одной и той же плоскости Oxy кривые

$$z_1 = f(x,y), \; z_2 = f(x,y),$$

так называемые линии уровня функции $f(x,y)$. По линиям уровня, если они соответствуют достаточно близким друг к другу значениям z, можно хорошо судить об изменениях функции $f(x, y)$, подобно тому как по линиям уровня топографической карты судят об изменениях рельефа местности.

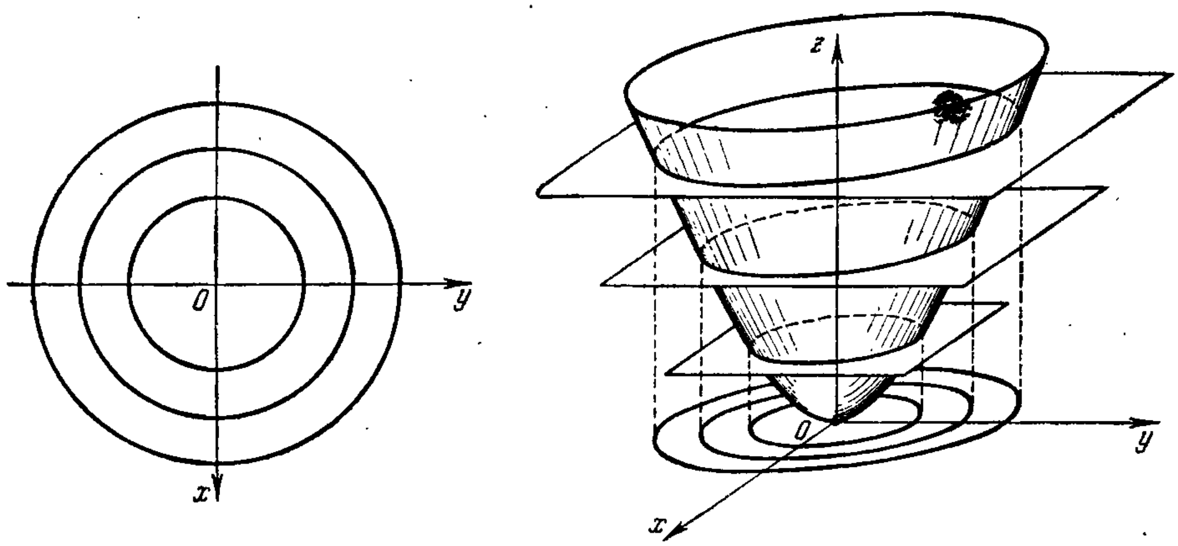


Рис. 27.

На рис. 27 изображена карта линий уровня функции $z = x^2 + y^2$; рядом показано, как они строятся. В главе III, на рис. 50, стр. 212 изображена подобная карта линий уровня для функции $z = xy$.

Частные производные и дифференциал. Сделаем несколько замечаний о дифференцировании функций нескольких переменных. Возьмем для примера какую-либо функцию

$$z = f(x, y)$$

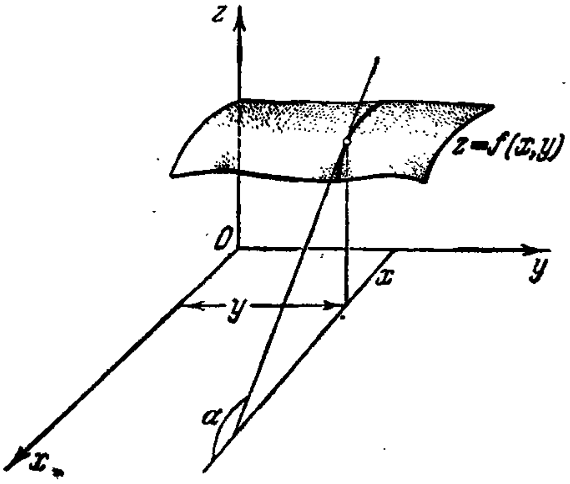


Рис. 28.

от двух переменных. Если зафиксировать значение y, т. е. считать его неизменяющимся, то наша функция от двух переменных превратится в функцию одной переменной x. Производная от нее, если она существует, называется *частной производной* по x и обозначается так:

$$\frac{\partial z}{\partial x}, \text{ или } \frac{\partial f}{\partial x}, \text{ или } f'_x(x, y).$$

Последнее обозначение подчеркивает, что частная производная по x есть, вообще говоря, функция от x и y. Подобным образом определяется и частная производная по y.

Геометрически наша функция в прямоугольной пространственной системе координат изображается поверхностью. Соответствующая функция от x при фиксированном y изображается плоской линией (рис. 28), полученной при сечении поверхности плоскостью, параллельной плоскости Oxz и отстоящей от этой последней на расстояние y. Частная производная $\frac{\partial z}{\partial x}$ равна, очевидно, тангенсу угла, образуемого касательной в точке (x, y) к этой линии с положительным направлением оси x-ов.

Вообще, если задана функция $z = f(x_1, x_2, \ldots, x_n)$ от переменных $x_1, x_2, \ldots, x_n$, то частной производной $\frac{\partial z}{\partial x_i}$ называется производная от нее по x_i, вычисленная при зафиксированных значениях прочих переменных:

$$x_1, x_2, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n.$$

Можно сказать, что частная производная от функции по переменной x_i есть скорость изменения этой функции в направлении изменения x_i. Можно было бы определить производную по произвольно заданному направлению, не обязательно совпадающему с той или иной осью координат, но мы на этом останавливаться не будем.

Примеры.

1) $z = \frac{x}{y}, \quad \frac{\partial z}{\partial x} = \frac{1}{y}, \quad \frac{\partial z}{\partial y} = -\frac{x}{y^2}.$

2) $u = \frac{1}{\sqrt{x^2 + y^2 + z^2}},$

$$\frac{\partial u}{\partial x} = -\frac{1}{x^2+y^2+z^2}\cdot\frac{2x}{2\sqrt{x^2+y^2+z^2}} = -\frac{x}{(x^2+y^2+z^2)^{3/2}}.$$

Иногда бывает необходимо от частных производных брать в свою очередь частные производные — это так называемые частные производные второго порядка. Для функции двух переменных их четыре

$$\frac{\partial^2 u}{\partial x^2},\quad \frac{\partial^2 u}{\partial x \partial y},\quad \frac{\partial^2 u}{\partial y \partial x},\quad \frac{\partial^2 u}{\partial y^2}.$$

Впрочем, в случае непрерывности этих производных вторая и третья из написанных (так называемые смешанные производные), как можно доказать, совпадают:

$$\frac{\partial^2 u}{\partial x\, \partial y} = \frac{\partial^2 u}{\partial y\, \partial x}.$$

Например, в случае первой рассмотренной выше функции

$$\frac{\partial^2 z}{\partial x^2} = 0,\qquad \frac{\partial^2 z}{\partial x \partial y} = -\frac{1}{y^2},\qquad \frac{\partial^2 z}{\partial y \partial x} = -\frac{1}{y^2},\qquad \frac{\partial^2 z}{\partial y^2} = \frac{2x}{y^3},$$

совпадение смешанных производных, как видит читатель, имеет место.

Для функций многих переменных, подобно тому как это было сделано для одной переменной, можно ввести понятие дифференциала.

Возьмем для определенности функцию

$$z = f(x, y)$$

от двух переменных. Если она имеет непрерывные частные производные, то можно доказать, что ее приращение

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

соответствующее приращениям Δx и Δy аргументов, можно представить в виде

$$\Delta z = \frac{\partial f}{\partial x}\Delta x + \frac{\partial f}{\partial y}\Delta y + \alpha\sqrt{\Delta x^2 + \Delta y^2},$$

где $\frac{\partial f}{\partial x}$ и $\frac{\partial f}{\partial y}$ — частные производные от функции в точке (x, y), а величина α зависит от Δx и Δy, и притом так, что $\alpha \to 0$, когда $\Delta x \to 0$ и $\Delta y \to 0$.

Сумма первых двух слагаемых

$$dz = \frac{\partial f}{\partial x}\Delta x + \frac{\partial f}{\partial y}\Delta y$$

линейно[1] зависит от Δx и Δy и называется *дифференциалом функции*.

[1] Вообще функция $Ax + By + C$, где A, B, C — постоянные, называется линейной функцией от x и y. Если $C = 0$, то она называется линейной однородной функцией. Здесь мы слово «однородная» опускаем.

Последнее слагаемое, благодаря наличию в нем множителя α, стремящегося к нулю вместе с Δx и Δy, есть бесконечно малая высшего порядка по отношению к величине

$$\rho = \sqrt{\Delta x^2 + \Delta y^2},$$

характеризующей общее изменение x и y.

Вот пример применения понятия дифференциала. Период колебания маятника вычисляется по формуле

$$T = 2\pi \sqrt{\frac{l}{g}},$$

где l — его длина, а g — ускорение силы тяжести. Предположим, что l и g нам известны с погрешностями, соответственно равными Δl и Δg. Тогда погрешность, с которой мы вычислим T, будет равна приращению ΔT, соответствующему приращениям аргументов Δl и Δg. Заменяя ΔT приближенно на dT, будем иметь

$$\Delta T \approx dT = \pi\left(\frac{\Delta l}{\sqrt{lg}} - \frac{\sqrt{l}\,\Delta g}{\sqrt{g^3}}\right).$$

Нам не известны знаки Δl и Δg, однако можно заведомо оценить ΔT неравенством

$$|\Delta T| \leqslant \pi\left(\frac{|\Delta l|}{\sqrt{lg}} + \sqrt{\frac{l}{g^3}}\,|\Delta g|\right),$$

из которого после деления на T получим

$$\frac{|\Delta T|}{T} \leqslant \left(\frac{|\Delta l|}{l} + \frac{|\Delta g|}{g}\right).$$

Итак, можно практически считать, что относительная ошибка для T равна сумме относительных ошибок для l и g.

В целях симметрии обозначений, приращения независимых переменных Δx и Δy принято обозначать символами dx и dy и также называть их дифференциалами. При таком обозначении дифференциал функции $u = f(x, y, z)$ запишется так:

$$du = \frac{\partial f}{\partial x}dx + \frac{\partial f}{\partial y}dy + \frac{\partial f}{\partial z}dz.$$

Частные производные играют большую роль всякий раз, когда приходится иметь дело с функциями многих переменных (а это бывает в огромном количестве приложений анализа к задачам техники и физики). С проблемой восстановления функции по свойствам ее частных производных мы еще встретимся в главе VI (том 2).

Ниже мы дадим простейшие примеры применения частных производных в анализе.

Дифференцирование неявных функций. Допустим, что нам надо узнать производную функции y от x, когда функция задана неявно зависимостью

$$F(x, y) = 0 \qquad (38)$$

между этими переменными. Если x, y удовлетворяют равенству (38) и мы дадим x приращение Δx, то y получит приращение Δy, такое, что $x+\Delta x$ и $y+\Delta y$ вновь удовлетворяют (38). Следовательно[1],

$$F(x+\Delta x,\ y+\Delta y)-F(x,y)=\frac{\partial F}{\partial x}\Delta x+\frac{\partial F}{\partial y}\Delta y+\alpha\sqrt{\Delta x^2+\Delta y^2}=0.$$

Отсюда, если только $\frac{\partial F}{\partial y}\neq 0$, уже следует

$$\lim_{\Delta x\to 0}\frac{\Delta y}{\Delta x}=y'_x=-\frac{\frac{\partial F}{\partial x}}{\frac{\partial F}{\partial y}}.$$

Мы получили способ находить производную неявно заданной функции y от x, не решая предварительно уравнения (38) относительно y.

Задачи на максимум и минимум. Если функция, скажем, двух переменных $z=f(x,y)$, достигает в точке $(x_0,\ y_0)$ максимума, т. е. если $f(x_0,\ y_0)\geqslant f(x,\ y)$ для всех точек $(x,\ y)$, близких к $(x_0,\ y_0)$, то эта точка должна также быть точкой максимального подъема для линии, полученной в сечении поверхности $z=f(x,y)$ плоскостью, параллельной Oxz или Oyz. Поэтому в такой точке обязательно выполняются условия

$$f'_x(x,y)=0,\ f'_y(x,\ y)=0. \tag{39}$$

Те же условия должны выполняться и в точке местного минимума. Поэтому наибольшее и наименьшее значения функции следует искать прежде всего среди точек, в которых выполняются условия (39). (Кроме того, не следует забывать о точках на границе области задания функции и точках, где функция не имеет производной, если такие точки есть.)

Чтобы установить, является ли найденная точка $(x,\ y)$, удовлетворяющая условиям (39), действительно точкой максимума или минимума, чаще всего пользуются различными косвенными соображениями. Например, если почему-либо известно, что функция дифференцируема и достигает минимума внутри области, а точек, где выполнены условия (39), там всего одна, то, очевидно, минимум именно в этой точке и достигается.

Пусть, например, требуется изготовить из жести прямоугольную коробку (без крышки) с данным объемом V, затратив возможно меньше материала. Если стороны основания этой коробки обозначить через x и y, то высота ее h будет равна V/xy, и, следовательно, поверхность S будет выражаться функцией

$$S=xy+\frac{V}{xy}(2x+2y)=xy+2V\left(\frac{1}{x}+\frac{1}{y}\right) \tag{40}$$

от x и y. Так как x и y по ходу задачи должны быть положительными, то вопрос свелся к отысканию минимума функции $S(x,y)$ среди всевозможных точек (x,y), заполняющих первую четверть координатной плоскости $(x,\ y)$, которую мы обозначим буквой G.

[1] Мы предполагаем, что $F(x,\ y)$ имеет непрерывные производные по x и y.

Если минимум достигается в какой-либо точке области G, то в ней должны равняться нулю частные производные

$$\frac{\partial S}{\partial x} = y - \frac{2V}{x^2} = 0,$$

$$\frac{\partial S}{\partial y} = x - \frac{2V}{y^2} = 0,$$

т. е. $yx^2 = 2V$, $xy^2 = 2V$, откуда и находим размеры коробки:

$$x = y = \sqrt[3]{2V} \text{ и } h = \sqrt[3]{\frac{V}{4}}. \tag{41}$$

Мы поставленную задачу решили, однако не совсем обосновали решение. Строгий математик скажет нам: «Вы предположили с самого начала, что при заданных условиях коробка с минимальной поверхностью существует и, исходя из этого допущения, нашли ее размеры. Таким образом, пока вы собственно получили только такое утверждение: если в G существует точка (x, y), для которой функция S достигает своего минимума, то ее координаты необходимо должны определяться равенствами (41). Докажите же, что минимум S на G существует, и тогда я признаю правильность вашего результата». Замечание это достаточно резонно, так как наша функция S, например, как мы скоро это увидим, максимума на области G не имеет. Покажем же, как можно убедиться, что в данном случае наша функция действительно достигает в некоторой точке (x, y) области G своего минимума.

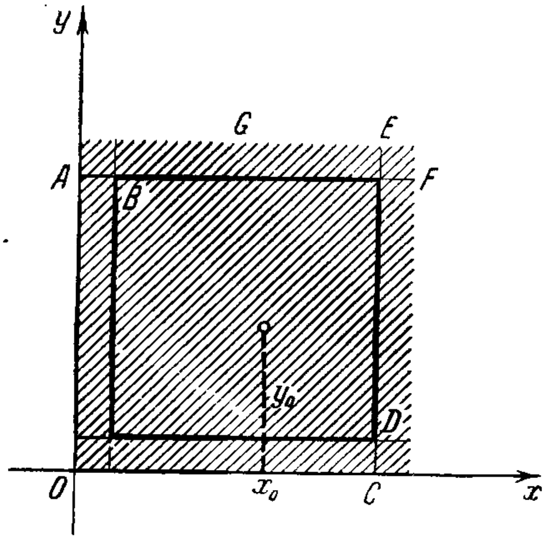


Рис. 29.

Основное утверждение, на котором мы при этом будем базироваться и которое доказывается совершенно строго в анализе, заключается в следующем. Если функция f от одной или нескольких переменных непрерывна всюду на некоторой конечной области H, ограниченной и содержащей свою границу, то в H всегда существует хоть одна точка, в которой эта функция достигает минимума (максимума). При помощи этого утверждения нам уже не будет трудно до конца разобраться в нашем примере.

Зададим любую точку (x_0, y_0) области G; пусть в ней $S(x_0, y_0) = N$. Зададим, далее, число R, одновременно удовлетворяющее неравенствам $R > N$, $2VR > N$, и построим квадрат Ω_R со стороной R^2, как на рис. 29, где $AB = CD = \frac{1}{R}$.

Оценим нашу функцию $S(x, y)$ снизу в точках области G, расположенных вне квадрата Ω_R. Если точка области G имеет абсциссу $x < \frac{1}{R}$, то

$$S(x, y) = xy + 2V\left(\frac{1}{x} + \frac{1}{y}\right) > 2V\frac{1}{x} > 2VR > N.$$

Аналогично, если точка области G имеет ординату $y < \frac{1}{R}$, то также $S > N$. Далее, если точка области G имеет абсциссу $x > \frac{1}{R}$ и расположена выше прямой AF или имеет ординату $y > \frac{1}{R}$ и расположена правее прямой CE, то

$$S(x, y) > xy > \frac{1}{R} R^2 = R > N.$$

Таким образом, для всех точек (x, y) области G, находящихся вне квадрата Ω_R, имеет место неравенство $S(x, y) > N$, и так как $S(x_0, y_0) = N$, то точка (x_0, y_0) принадлежит квадрату и, следовательно, минимум нашей функции на G равен минимуму ее на квадрате.

Но функция $S(x, y)$ непрерывна внутри этого квадрата и на его границе, поэтому на основании сформулированного выше утверждения существует в квадрате точка (x, y), где наша функция достигает минимума по отношению к квадрату, а следовательно, и в нашей области G. Этим существование минимума доказано.

Приведенные рассуждения служат примером того, как можно рассуждать при отыскании максимума или минимума функции, заданной на неограниченной области.

Формула Тейлора. Функции многих переменных, так же как функции одной переменной, могут представляться по формуле Тейлора. Например, разложение функции

$$u = f(x, y)$$

в окрестности точки (x_0, y_0), если ограничиться первыми и вторыми степенями $x - x_0$ и $y - y_0$, имеет вид

$$\begin{aligned} f(x, y) = f(x_0, y_0) &+ [f'_x(x_0, y_0)(x - x_0) + f'_y(x_0, y_0)(y - y_0)] + \\ &+ \frac{1}{2!}[f''_{xx}(x_0, y_0)(x - x_0)^2 + 2f''_{xy}(x_0, y_0)(x - x_0)(y - y_0) + \\ &+ f''_{yy}(x_0, y_0)(y - y_0)^2] + R_3. \end{aligned}$$

При этом, если функция $f(x, y)$ имеет непрерывные частные производные 2-го порядка, то остаточный член стремится к нулю быстрее, чем

$$r^2 = (x - x_0)^2 + (y - y_0)^2,$$

т. е. быстрее, чем квадрат расстояния между точками (x, y) и (x_0, y_0), когда $r \to 0$. Формула Тейлора дает весьма общее средство задания и приближенного вычисления значений различных функций.

Отметим, что с помощью этой формулы можно также решать поставленный выше вопрос о том, имеет ли функция максимум или минимум в точке, где $\frac{\partial f}{\partial x} = \frac{\partial f}{\partial y} = 0$. Действительно, если эти условия выполнены в некоторой точке (x_0, y_0), то для точек (x, y), близких к (x_0, y_0), значение функции будет, согласно формуле Тейлора, отличаться от $f(x_0, y_0)$ на величину

$$f(x, y) - f(x_0, y_0) =$$
$$= \frac{1}{2!}[A(x-x_0)^2 + 2B(x-x_0)(y-y_0) + C(y-y_0)^2] + R_3, \qquad (42)$$

где A, B и C соответственно обозначают вторые частные производные f''_{xx}, f''_{xy}, f''_{yy} в точке (x_0, y_0).

Если окажется, что функция

$$\Phi(x, y) = A(x-x_0)^2 + 2B(x-x_0)(y-y_0) + C(y-y_0)^2$$

при любых, не равных одновременно нулю $(x - x_0)$ и $(y - y_0)$ положительна, то в таком случае вся правая часть равенства (42) при малых $(x - x_0)$ и $(y - y_0)$ будет положительной, так как при достаточно малых $(x - x_0)$ и $(y - y_0)$ величина R_3 по абсолютной величине заведомо меньше $\frac{1}{2}\Phi(x,y)$. Отсюда будет следовать тогда, что в точке (x_0, y_0) функция f достигает минимума. Наоборот, если функция $\Phi(x, y)$ при любых $(x - x_0)$ и $(y - y_0)$ отрицательна, то это влечет за собой отрицательность всей правой части равенства (42) при малых $(x - x_0)$ и $(y - y_0)$, и в точке (x_0, y_0) имеет место максимум.

В более сложных случаях приходится рассматривать следующие члены формулы Тейлора.

Задачи о максимуме и минимуме функций трех и более переменных решаются и исследуются совершенно аналогичным путем. В виде упражнения читатель может доказать, что если в пространстве в заданных точках

$$P_1(x_1, y_1, z_1), P_2(x_2, y_2, z_2), \ldots, P_n(x_n, y_n, z_n)$$

расположены данные массы

$$m_1, m_2, \ldots, m_n,$$

то момент M этой системы относительно точки $P(x, y, z)$, равный сумме произведений масс на квадраты их расстояний до точки P

$$M(x, y, z) = \sum_{i=1}^{n} m_i [(x-x_i)^2 + (y-y_i)^2 + (z-z_i)^2],$$

будет минимальным из возможных, если точку P поместить в так называемом центре тяжести системы, имеющем координаты

$$x = \frac{\sum_{i=1}^{n} m_i x_i}{\sum_{i=1}^{n} m_i}, \quad y = \frac{\sum_{i=1}^{n} m_i y_i}{\sum_{i=1}^{n} m_i}, \quad z = \frac{\sum_{i=1}^{n} m_i z_i}{\sum_{i=1}^{n} m_i}.$$

Относительный максимум и минимум. Для функций многих переменных можно ставить несколько измененные задачи на максимум и минимум. Поясним это на простом примере. Допустим, что среди всех прямоугольников, вписанных в окружность радиуса R, мы хотим найти прямоугольник, имеющий наибольшую площадь. Площадь прямоугольника равна произведению xy его сторон, причем x и y положительны и связаны в данном случае соотношением $x^2+y^2=(2R)^2$, как это ясно из рис. 30. Итак, требуется найти максимум функции $f(x, y)=xy$ среди только таких x и y, которые удовлетворяют соотношению $x^2+y^2=4R^2$.

Подобные задачи, в которых надо отыскать максимум (или минимум) некоторой функции $f(x, y)$ только среди тех x, y, которые связаны между собой некоторым соотношением $\varphi(x, y)=0$, на практике возникают весьма часто.

Разумеется, можно решить уравнение $\varphi(x, y)=0$ относительно y, полученное выражение y подставить в $f(x, y)$ и искать обычный максимум уже для функции одной переменной x. Но этот путь обычно сложен, а иногда и невыполним.

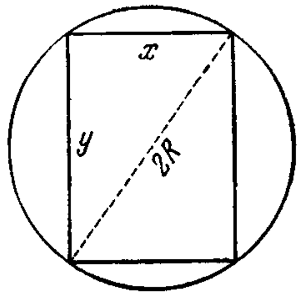


Рис. 30.

Для решения подобных задач в анализе выработан несравненно более удобный прием, называемый методом множителей Лагранжа. Идея его совсем проста. Рассмотрим функцию

$$F(x, y)=f(x, y)+\lambda\varphi(x, y),$$

где λ — произвольное постоянное число. Очевидно для x, y, удовлетворяющих условию $\varphi(x, y)=0$, значения $F(x, y)$ совпадают с $f(x, y)$.

Для функции $F(x, y)$ будем искать максимум, не связывая x, y никакими ограничениями. В точке максимума должны быть соблюдены условия $\frac{\partial F}{\partial x}=\frac{\partial F}{\partial y}=0$[1], или иначе

$$\frac{\partial f}{\partial x}+\lambda\frac{\partial \varphi}{\partial x}=0; \tag{43}$$

$$\frac{\partial f}{\partial y}+\lambda\frac{\partial \varphi}{\partial y}=0. \tag{44}$$

Значения x и y в точке максимума $F(x, y)$ служат одним из решений системы уравнений (43), (44) и зависят, естественно, от коэффициента λ, входящего в эти уравнения. Допустим теперь, что нам удалось подобрать число λ так, что координаты точки максимума удовлетворяют условию

$$\varphi(x, y)=0. \tag{45}$$

Тогда эта точка будет и точкой местного максимума в исходной задаче.

---

[1] Речь идет, конечно, о максимуме, достигаемом внутри области задания функции $F(x, y)$. Функции $f(x, y)$ и $\varphi(x, y)$ предполагаются дифференцируемыми.

В самом деле, на нашу задачу можно геометрически смотреть так. Функция $f(x,y)$ задана в некоторой области G (рис. 31). Условию $\varphi(x, y) = 0$ обычно удовлетворяют точки некоторой кривой Г. Надо найти наибольшее значение $f(x, y)$ в точках линии Г. Если в точке на кривой Г достигнут максимум $F(x, y)$, то $F(x, y)$ не возрастает при небольших сдвигах в любую сторону из этой точки, в том числе при сдвигах по кривой Г. Но при сдвигах вдоль Г значения $F(x, y)$ совпадают с $f(x, y)$, значит при малых сдвигах по кривой не возрастает функция $f(x,y)$, и она имеет в этой точке местный максимум.

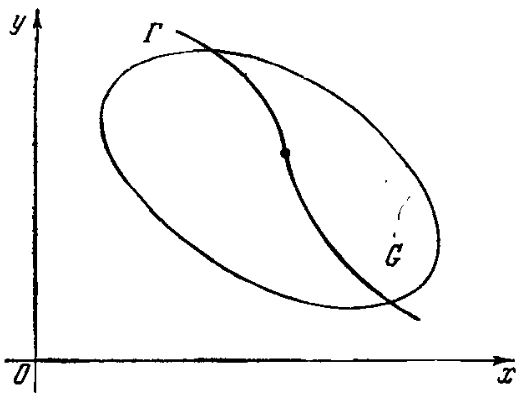


Рис. 31.

Эти соображения подсказывают простой прием решения задачи. Составляем уравнения (43), (44), (45); решаем эту систему уравнений относительно неизвестных x, y, λ; получится одно или несколько решений

$$(x_1, y_1, \lambda_1),\ (x_2, y_2, \lambda_2), \ldots \tag{46}$$

Присоединим к точкам (x_1, y_1), $(x_2, y_2), \ldots$ еще точки границы G, где кривая Г выходит из области G, после чего выбираем из всех этих точек ту, в которой $f(x, y)$ принимает наибольшее (или наименьшее) значение.

Конечно, наши наводящие рассуждения еще вовсе не доказывают правильности этого приема. В самом деле, ведь не доказано, что точка, где $f(x, y)$ достигает местного максимума по отношению к соседним точкам кривой Г, получается как точка максимума функции $F(x, y)$ при каком-нибудь λ. Можно, однако, показать, — и это делается во всех учебниках анализа, — что всякая точка (x_0, y_0) местного максимума $f(x, y)$ на кривой Г будет указанным путем обнаружена, если только в этой точке не обращаются одновременно в нуль производные[1] $\varphi'_x(x_0, y_0)$, $\varphi'_y(x_0, y_0)$.

Решим методом Лагранжа пример, приведенный в начале настоящего пункта. В этом случае $f(x, y) = xy$; $\varphi(x, y) = x^2 + y^2 - 4R^2$. Составляем уравнения (43), (44), (45)

$$\begin{aligned} y + 2\lambda x &= 0, \\ x + 2\lambda y &= 0, \\ x^2 + y^2 &= 4R^2, \end{aligned}$$

[1] В курсе высшей математики В. И. Смирнова (т. I, стр. 394) читатель может найти простой пример, когда последняя особенность приводит к потере решения, если мы будем механически применять метод Лагранжа и не рассмотрим в дополнение к перечисленным выше точку, в которой имеет место наряду с (45) условие:

$$\varphi'_x(x_0, y_0) = 0, \quad \varphi'_y(x_0, y_0) = 0.$$

откуда, учитывая, что x и y положительны, находим единственное решение

$$x = y = R\sqrt{2} \quad \left(\lambda = -\frac{1}{2}\right).$$

Для таких равных друг другу x и y, т. е. в случае вписанного квадрата, действительно достигается максимум площади.

Метод Лагранжа распространяется и на случай функций от трех и более переменных. Дополнительных условий, подобных условию (45), при этом может быть несколько (но меньше, чем переменных), и тогда вводится соответствующее количество вспомогательных множителей.

Вот еще примеры задач на отыскание относительного максимума или минимума.

Пример 1. При какой высоте h и каком радиусе r открытый цилиндрический бак данной вместимости V потребует для своего изготовления меньше листового материала, т. е. будет иметь меньшую площадь стенок и круглого днища?

Задача, очевидно, сводится к тому, чтобы найти минимум функции переменных r и h:

$$f(r, h) = 2\pi rh + \pi r^2$$

при соблюдении условия $\pi r^2 h = V$, которое можно записать в виде

$$\varphi(r, h) = \pi r^2 h - V = 0.$$

Пример 2. Движущаяся точка должна пройти из A в B (рис. 32). На пути AM она движется со скоростью v_1, а на пути MB — со скоростью v_2. Где на прямой DD' надо поместить точку M, чтобы весь путь из A в B был пройден возможно быстрее?

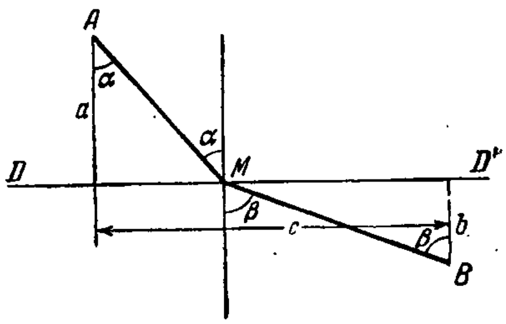


Рис. 32.

Примем углы α и β, обозначенные на рис. 32, за неизвестные. Длины a и b перпендикуляров из точек A и B на прямую DD' и расстояние c между основаниями этих перпендикуляров нам известны. Время прохождения всего пути выразится, как легко видеть, формулой

$$f(\alpha, \beta) = \frac{a}{v_1 \cos\alpha} + \frac{b}{v_2 \cos\beta}.$$

Надо найти минимум этого выражения, учитывая, что α и β связаны соотношением

$$a \operatorname{tg}\alpha + b \operatorname{tg}\beta = c.$$

Приведенные примеры читатель может решить сам, используя метод Лагранжа. Во втором примере легко убедиться, что при наивыгоднейшем расположении точки M выполняется условие

$$\frac{\sin\alpha}{\sin\beta} = \frac{v_1}{v_2}.$$

Это — известный закон преломления светового луча. Таким образом, световой луч преломляется при переходе из одной среды в другую так, что время его прохождения из точки одной среды в точку другой минимально. Такого рода выводы представляют уже не только вычислительный, но и большой познавательный интерес; они побуждают точное естествознание к проникновению в более глубокие и общие закономерности природы.

Наконец, отметим, что множители λ, вводимые при решении задач методом Лагранжа, не остаются только вспомогательными числами. Они оказываются всякий раз тесно связанными с существом частной задачи и приобретают в связи с ней конкретный смысл.

§ 13. ОБОБЩЕНИЯ ПОНЯТИЯ ИНТЕГРАЛА

В § 10 мы называли определенным интегралом от функции $f(x)$ на отрезке $[a, b]$ предел сумм

$$\sum_{i=1}^{n} f(\xi_i)\,\Delta x_i$$

при стремлении к нулю наибольшего участка Δx_i разбиения отрезка $[a, b]$. Несмотря на то, что класс функций $f(x)$, для которых этот предел действительно существует (класс интегрируемых функций), весьма широк, в частности охватывает все непрерывные и даже многие разрывные функции, этот класс функций обладает серьезным недостатком. Складывая, вычитая, умножая, а при определенных условиях и деля значения двух интегрируемых функций $f(x)$ и $\varphi(x)$, мы получаем, как это можно доказать, функции, снова интегрируемые. Для $\frac{f(x)}{\varphi(x)}$ это во всяком случае верно, если величина $\frac{1}{\varphi(x)}$ остается ограниченной на $[a, b]$. Но если мы некоторую функцию получили в результате предельного перехода от приближающих ее интегрируемых функций $f_1(x)$, $f_2(x)$, $f_3(x)$, ... так, что при всех x из $[a, b]$

$$f(x) = \lim_{n\to\infty} f_n(x),$$

то предельная функция $f(x)$ не обязательно окажется интегрируемой.

В ряде случаев это и другие обстоятельства создают большие осложнения в самом математическом аппарате, широко пользующемся процессами предельного перехода.

Выход из положения был найден в дальнейших обобщениях понятия интеграла. Важнейшим из таких обобщений является интеграл Лебега, с которым читатель познакомится в главе XV (том 3), посвященной теории функций действительного переменного. Мы же должны будем здесь остановиться на весьма важных для практики обобщениях интеграла в другом направлении.

Кратные интегралы. Мы познакомились с процессом интегрирования функции одной переменной, заданной на одномерной области — отрезке. Но аналогичный процесс возможно распространить также и на функции двух, трех и вообще многих переменных, заданные на соответствующих областях.

Пусть, например, в прямоугольной системе координат задана поверхность

$$z = f(x, y),$$

а в плоскости Oxy задана область G, ограниченная замкнутой кривой Γ. Требуется определить объем, ограниченный нашей поверхностью, плоскостью Oxy и цилиндрической поверхностью, проходящей через кривую Γ с образующей, параллельной оси Oz (рис. 33). Чтобы решить эту задачу, разделим плоскую область G на частичные области при помощи какой-либо сетки прямых, параллельных осям Ox и Oy, и перенумеруем те из этих частичных областей

$$G_1, G_2, \ldots, G_n,$$

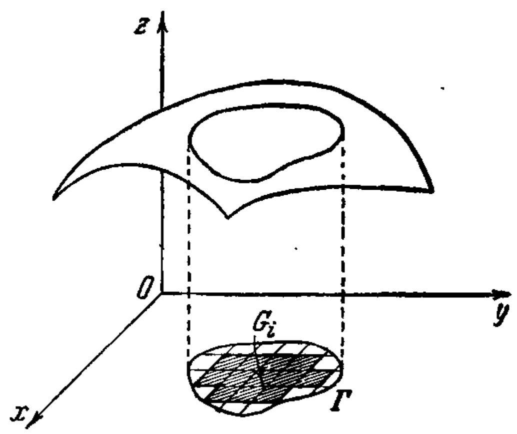


Рис. 33.

которые представляют собой полные прямоугольники. Если сетка достаточно густа, то значительная часть нашей области G будет исчерпана перенумерованными прямоугольниками. В каждом из них мы выберем по произвольной точке

$$(\xi_1, \eta_1), (\xi_2, \eta_2), \ldots, (\xi_n, \eta_n)$$

и, считая для простоты, что G_i обозначает не только прямоугольник, но и число, равное его площади, составим сумму

$$S_n = f(\xi_1, \eta_1) G_1 + f(\xi_2, \eta_2) G_2 + \ldots + f(\xi_n, \eta_n) G_n = \sum_{i=1}^{n} f(\xi_i, \eta_i) G_i. \tag{47}$$

Очевидно, если поверхность непрерывна, а сетка достаточно густа, то эта сумма может быть сделана как угодно близкой к искомому объему V. Мы получим наш объем точно, если возьмем предел суммы (47) при все более мелких разбиениях (т. е. в предположении, что даже наибольшая из диагоналей наших прямоугольников стремится к нулю)

$$\lim_{\max d(G_i) \to 0} \sum_{i=1}^{n} f(\xi_i, \eta_i) G_i = V. \tag{48}$$

С аналитической точки зрения для определения объема V потребовалось произвести над функцией $f(x, y)$ и областью G, где она задана, некоторую математическую операцию, указанную в левой части равенства (48). Эта операция называется операцией интегрирования f в области G, а результат ее — интегралом от функции f по области G. Этот результат принято обозначать следующим образом:

$$\iint_G f(x, y)\,dx\,dy = \lim_{\max d(G_i)\to 0} \sum_{i=1}^{n} f(\xi_i, \eta_i)\,G_i. \tag{49}$$

Подобным образом можно определить интеграл от функции трех переменных по трехмерной области G, представляющей собой некоторое тело в пространстве. Мы снова делим область G на части на этот раз плоскостями, параллельными координатным плоскостям в пространстве. Выбираем среди этих частичных областей полные параллелепипеды и перенумеровываем их

$$G_1, G_2, \ldots, G_n.$$

В каждом из них берем по произвольной точке

$$(\xi_1, \eta_1, \zeta_1), (\xi_2, \eta_2, \zeta_2), \ldots, (\xi_n, \eta_n, \zeta_n)$$

и составляем сумму

$$S = \sum_{i=1}^{n} f(\xi_i, \eta_i, \zeta_i)\,G_i, \tag{50}$$

где G_i обозначает объем параллелепипеда G_i. Наконец, интеграл от $f(x, y, z)$ по области G определяем как предел

$$\lim_{\max d(G_i)\to 0} \sum_{i=1}^{n} f(\xi_i, \eta_i, \zeta_i)\,G_i = \iiint_G f(x, y, z)\,dx\,dy\,dz, \tag{51}$$

к которому стремится сумма (50), когда максимальная диагональ $d(G_i)$ стремится к нулю.

Рассмотрим пример. Представим себе, что область G заполнена неравномерно массой, причем известна функция $\rho(x, y, z)$, выражающая плотность распределения массы в G. Плотность $\rho(x, y, z)$ массы в точке (x, y, z) определяется как предел, к которому стремится отношение массы какой-либо малой области, содержащей точку (x, y, z), к ее объему, когда диаметр[1] этой области стремится к нулю. Чтобы определить массу тела G, естественно рассуждать так. Делим область G на части плоскостями, параллельными плоскостям координат, перенумеровываем образованные при этом полные параллелепипеды

$$G_1, G_2, \ldots, G_n.$$

Если плоскости, делящие G, расположены достаточно густо, то мы сделаем малую ошибку, если пренебрежем массами неправильных частичных областей, а массу каждой правильной области G_i (полного параллелепипеда) определим приближенно как произведение

$$\rho(\xi_i, \eta_i, \zeta_i)\,G_i,$$

[1] Диаметром области называется верхняя грань расстояний между двумя точками области.

где (ξ_i, η_i, ζ_i) — какая-либо точка G_i. В результате приближенное значение массы M будет выражаться суммой

$$S_n = \sum_{i=1}^{n} \rho(\xi_i, \eta_i, \zeta_i) G_i,$$

а точное ее выражение, очевидно, есть предел этой суммы, когда максимальная диагональ G_i стремится к нулю, т. е.

$$M = \iiint_G \rho(x, y, z)\, dx\, dy\, dz = \lim_{\max d(G_i) \to 0} \sum_{i=1}^{n} \rho(\xi_i, \eta_i, \zeta_i) G_i.$$

Интегралы (49) и (51) носят соответственно название дву- и трехкратных интегралов.

Разберем задачу, приводящую к двукратному интегралу. Представим себе, что по плоской поверхности течет вода. Кроме того, на поверхность с разной интенсивностью $f(x, y)$ в разных местах просачивается подпочвенная вода (или, наоборот, вода уходит в грунт). Выделим ограниченную замкнутым контуром область G (рис. 34) и допустим, что нам известна интенсивность $f(x, y)$, т. е. количество выделяющейся в минуту подпочвенной воды на 1 *см*² поверхности, в каждой точке области G ($f(x, y) > 0$ там, где выделяется подпочвенная вода, и $f(x, y) < 0$ там, где вода просачивается в грунт). Сколько воды выделяется в минуту во всей области G?

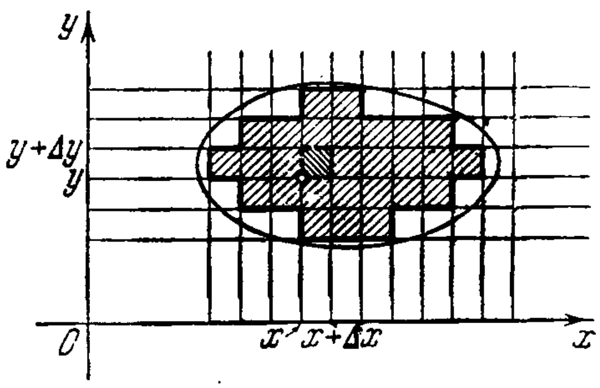


Рис. 34.

Если мы разобьем область G на небольшие участки и подсчитаем количество выступающей воды приближенно, считая на каждом участке $f(x, y)$ постоянной, а затем перейдем к пределу при все более мелких разбиениях, то получим выражение общего количества отданной грунтом воды в виде интеграла

$$\iint_G f(x, y)\, dx\, dy.$$

Двойные (двукратные) интегралы впервые ввел в рассмотрение Эйлер. Кратные интегралы служат повседневно употребляемым средством в самых разнообразных расчетах и исследованиях.

Можно было бы показать, но это не входит в нашу задачу, что вычисление кратных интегралов, как правило, может быть сведено к повторному вычислению обыкновенных одномерных интегралов.

Контурные и поверхностные интегралы. Наконец, следует отметить, что возможны еще другие обобщения интеграла. Так, например, задача об определении работы переменной силы, приложенной к материальной точке, движущейся по заданной кривой, естественно приводит к так называемому криволинейному интегралу, а задача о нахождении общего заряда поверхности, на которой непрерывно распределено электричество с данной поверхностной плотностью, приводит к новому понятию — интеграла по поверхности.

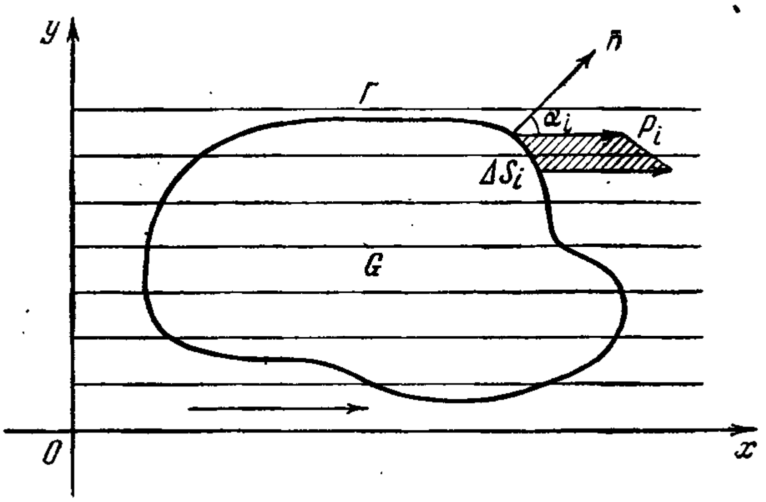


Рис. 35.

Пусть, например, в пространстве течет жидкость (рис. 35) и при этом скорость частицы жидкости в точке (x, y) выражается функцией $P(x, y)$. Таким образом скорость не зависит от z. Если мы хотим определить количество жидкости, вытекающей в минуту за пределы контура Г[1], то можно рассуждать следующим образом. Разобьем Г на участки Δs_i. Количество воды, протекающей через один участок Δs_i, приближенно равно заштрихованному на рис. 35 столбику жидкости, как бы выдавливающемуся за минуту через этот участок контура. Но площадь заштрихованного параллелограма равна

$$P_i(x, y)\cdot\Delta s_i\cdot\cos\alpha_i,$$

где α_i — угол между направлением $\bar{x}$ оси x-ов и идущим наружу из области, ограниченной контуром Г, перпендикуляром $\bar{n}$ к направлению касательной, вдоль которой можно приближенно считать направленным отрезок Δs_i. Суммируя площади таких параллелограмов и переходя к пределу при все более мелких разбиениях контура Г, получим количество вытекающей в минуту через контур Г воды, которое обозначают так:

$$\int_{\Gamma} P(x, y)\cos(\bar{n}, \bar{x})\,ds$$

и называют криволинейным интегралом. Если поток не будет параллельным, то скорость течения будет иметь в каждой точке (x, y) некоторые составляющие $P(x, y)$ вдоль оси x-ов и $Q(x, y)$ вдоль оси y-ов. В этом случае аналогичным рассуждением можно установить, что количество вытекаю-

[1] Точнее, через цилиндрическую поверхность, имеющую основанием контур Г и высоту равную единице.

щей через контур Г воды будет равно

$$\int_{\Gamma}[P(x,\ y)\cos(\overline{n},\ \overline{x})+Q(x,\ y)\cos(\overline{n},\ \overline{y})]\,ds$$[1].

Когда речь идет об интеграле по кривой поверхности G от заданной в ее точках $M(x, y, z)$ функции $f(M)$, то подразумевается предел сумм вида

$$\lim\sum_{i=1}^{n} f(M_i)\,\Delta\sigma_i=\iint_{G} f(x,\ y,\ z)\,d\sigma$$

при все более мелких разбиениях области G, по которой производится интегрирование, на участки, площади которых равны $\Delta\sigma_i$.

Для кратных, криволинейных и поверхностных интегралов существуют как общие методы их вычисления и преобразования, так и приближенные методы.

Формула Остроградского. Очень важные зависимости между интегралом по некоторому объему (или области) и интегралом по поверхности этого объема (или границе области) были в весьма общем виде обнаружены в середине прошлого столетия М. В. Остроградским.

Не ставя себе целью доказать здесь общую формулу Остроградского, имеющую самое широкое применение, попробуем на примере пояснить эту формулу в ее наиболее простом частном случае.

Представим себе, что по плоской поверхности грунта течет установившийся поток воды, причем часть воды все время просачивается в грунт или поступает из грунта. Выделяем некоторую область G, ограниченную контуром Г, и предполагаем известными для каждой точки этой области составляющие $P(x, y)$ и $Q(x, y)$ скорости воды в направлениях оси x-ов и оси y-ов.

Подсчитаем, с какой интенсивностью вблизи точки с координатами (x, y) вода поступает из грунта. Для этого рассмотрим маленький прямоугольник со сторонами Δx и Δy, прилегающий к точке (x, y).

Благодаря составляющей скорости $P(x, y)$ через левую вертикальную границу этого участка в него втекает за минуту приближенно $P(x, y)\,\Delta y$ единиц воды, а через правую границу за то же время вытекает приближенно $P(x+\Delta x, y)\,\Delta y$ единиц воды. В общем на единицу площади через левую и правую вертикальные границы квадратика вытекает приближенно количество воды, равное

$$\frac{[P(x+\Delta x,\ y)-P(x,\ y)]\,\Delta y}{\Delta x\,\Delta y}.$$

---

[1] Поскольку при малых смещениях по кривой дифференциал координаты y равен как раз $\cos(\overline{n},\ \overline{x})\,ds$, а дифференциал dx равен $-\cos(\overline{n},\ \overline{y})\,ds$, то последний интеграл часто записывают в виде

$$\int_{\Gamma}[P(x,\ y)\,dy-Q(x,\ y)\,dx].$$

Если заставить Δx стремиться к нулю, то получим в пределе

$$\frac{\partial P}{\partial x}.$$

Соответственно интенсивность, с которой вода вытекает из малого участка в направлении оси y-ов, выразится величиной

$$\frac{\partial Q}{\partial y}.$$

Значит, интенсивность выделения грунтовой воды в точке с координатами (x, y) будет равна

$$\frac{\partial P}{\partial x}+\frac{\partial Q}{\partial y}.$$

Общее же количество поступающей из грунта воды равняется, как мы уже убедились раньше, двойному интегралу от функции, выражающей интенсивность выделения грунтовой воды в каждой точке, т. е.

$$\iint\limits_{G}\left(\frac{\partial P}{\partial x}+\frac{\partial Q}{\partial y}\right)dx\,dy. \tag{52}$$

Все это количество воды должно вытекать за то же время из пределов контура Г. Но количество воды, вытекающей через контур Г, как мы видели, выражается криволинейным интегралом по Г

$$\int\limits_{\Gamma}[P(x, y)\cos(\overline{n}, \overline{x})+Q(x, y)\cos(\overline{n}, \overline{y})]\,ds. \tag{53}$$

Равенство величин (52) и (53) и выражает собой формула Остроградского в ее простейшем двухмерном случае

$$\iint\limits_{G}\left(\frac{\partial P}{\partial x}+\frac{\partial Q}{\partial y}\right)dx\,dy=\int\limits_{\Gamma}[P(x, y)\cos(\overline{n}, \overline{x})+Q(x, y)\cos(\overline{n}, \overline{y})]\,ds.$$

Мы только пояснили смысл этой формулы на физическом примере. Она может быть доказана математически.

Таким образом, математическая теорема Остроградского отражает определенную закономерность действительности, которую мы в нашем примере наглядно воспринимаем как баланс сохранения количества несжимаемой жидкости.

М. В. Остроградский доказал значительно более общую формулу связи интеграла по многомерному объему с интегралом по его границе. В частности, для трехмерного тела G, ограниченного поверхностью Г, эта формула принимает вид

$$\iiint\limits_{G}\left(\frac{\partial P}{\partial x}+\frac{\partial Q}{\partial y}+\frac{\partial R}{\partial z}\right)dx\,dy\,dz=\iint\limits_{\Gamma}[P\cos(\overline{n}, \overline{x})+Q\cos(\overline{n}, \overline{y})+R\cos(\overline{n}, \overline{z})]\,d\sigma,$$

где $d\sigma$ элемент поверхности.

Интересно отметить, что основную теорему интегрального исчисления

$$\int_a^b f(x)\,dx = F(b) - F(a) \tag{54}$$

можно рассматривать как одномерный случай формулы Остроградского. Равенство (54) связывает интеграл по отрезку с «интегралом» по его «нульмерной» границе, состоящей из двух концевых точек.

Формулу (54) можно пояснить следующей аналогией. Вообразим, что в прямой трубе с постоянным сечением $s = 1$ течет вода со скоростью $F(x)$, различной в разных сечениях (рис. 36). Через пористые стенки трубы в нее (или из нее) с различной интенсивностью в разных сечениях

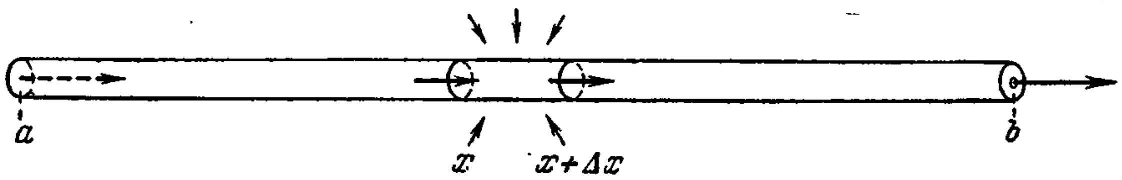


Рис. 36.

просачивается вода извне. Если рассмотреть участок от x до $x + \Delta x$, то количество воды, просачивающейся за единицу времени на этом участке в трубу, должно компенсировать разницу $F(x + \Delta x) - F(x)$ между количеством вытекающей и количеством втекающей в этот участок вдоль трубы воды. Поэтому количество просачивающейся на этом участке воды совпадает с разностью $F(x + \Delta x) - F(x)$, а интенсивность $f(x)$, выражающая отношение количества просачивающейся на бесконечно малом участке воды к длине этого участка, будет равна:

$$f(x) = \lim_{\Delta x \to 0} \frac{F(x + \Delta x) - F(x)}{\Delta x} = F'(x).$$

Общее количество воды, втекающей в трубу на участке $[a, b]$ и вытекающей из нее, должно быть одинаковым. Но через стенки втекает количество, равное $\int_a^b f(x)\,dx$, а вдоль трубы через границы участка вытекает количество $F(b) - F(a)$. Равенство этих величин и выражает как раз формула (54).

§ 14. РЯДЫ

Понятие ряда. *Рядом* в математике называют выражение вида

$$u_0 + u_1 + u_2 + \ldots$$

Числа u_k называются членами ряда. Они задаются в бесконечном числе и располагаются в определенном порядке, так что каждому натуральному числу $k = 0, 1, 2, \ldots$ соответствует определенное значение u_k.

Читатель должен иметь в виду, что мы пока не сказали, можно ли вычислять такие выражения и как это делать. То обстоятельство, что между членами u_k в нашем выражении, стоят знаки плюс, указывает как будто на то, что все члены надо сложить. Однако их бесконечно много, а сложение чисел определено только для конечного числа слагаемых.

Обозначим через S_n сумму первых n членов нашего ряда; ее называют *n-й частичной суммой* ряда. В результате мы получим последовательность чисел

$$\begin{aligned} &S_1 = u_0, \\ &S_2 = u_0 + u_1, \\ &\dots\dots\dots\dots\dots \\ &S_n = u_0 + u_1 + \dots u_{n-1}, \\ &\dots\dots\dots\dots\dots \end{aligned}$$

и можем говорить о переменной величине S_n, где $n = 1, 2, \dots$

Ряд называется *сходящимся*, если переменная S_n стремится при $n \to \infty$ к определенному конечному пределу

$$\lim_{n\to\infty} S_n = S.$$

Сам этот предел называется *суммой ряда*. В этом случае пишут

$$S = u_0 + u_1 + u_2 + \dots$$

Если же предела S_n при $n \to \infty$ не существует, то ряд называется *расходящимся*, и в этом случае не имеет смысла говорить о его сумме[1]. Впрочем, если все u_n одного знака, то в этом случае принято говорить, что сумма ряда равна бесконечности с соответствующим знаком.

Рассмотрим в качестве примера ряд

$$1 + x + x^2 + \dots,$$

члены которого образуют геометрическую прогрессию со знаменателем x.

Сумма его первых n членов равна

$$S_n(x) = \frac{1 - x^n}{1 - x} \qquad (x \neq 1); \qquad (55)$$

в случае, если $|x| < 1$, эта сумма имеет предел

$$\lim_{n\to\infty} S_n(x) = \frac{1}{1 - x},$$

и потому для $|x| < 1$ можно написать

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

---

[1] Заметим, что возможны обобщенные определения суммы ряда, в силу которых некоторым расходящимся рядам может быть приписано более или менее естественно понятие «обобщенной суммы». Такие ряды называют суммируемыми. Оперирование с обобщенными суммами расходящихся рядов иногда оказывается полезным.

Если $|x|>1$, то, очевидно,

$$\lim_{n\to\infty} S_n(x) = \infty,$$

и ряд расходится. Это же обстоятельство имеет место и при $x=1$, в чем можно убедиться непосредственно без формулы (55), которая при $x=1$ не имеет смысла. Наконец, при $x=-1$ сумма первых членов ряда принимает попеременно значения то $+1$, то 0, и ряд также будет расходящимся.

Каждому ряду соответствует определенная последовательность значений частичных его сумм $S_1, S_2, S_3, \ldots$, от стремления которых к пределу и зависит сходимость ряда. Наоборот, любой последовательности чисел $S_1, S_2, S_3, \ldots$ соответствует ряд

$$S_1 + (S_2 - S_1) + (S_3 - S_2) + \ldots,$$

частичными суммами которого будут числа этой последовательности. Таким образом, теория переменных, пробегающих последовательность, может быть сведена к теории соответствующих рядов, и наоборот. Однако каждая из этих теорий имеет самостоятельное значение. В одних случаях целесообразнее переменную изучать непосредственно, в других — рассматривать эквивалентный ей ряд.

Отметим, что ряды с давних пор служили важным средством для вычисления и представления различных величин и прежде всего функций. Разумеется, взгляды математиков на понятие ряда исторически менялись и находились в тесной связи с состоянием развития всего анализа бесконечно малых. То четкое определение сходимости и расходимости ряда, которое дано выше, сформировалось в начале прошлого столетия вместе с органически связанным с ним понятием предела.

Если ряд сходится, то его общий член стремится к нулю при неограниченном возрастании n, так как

$$\lim_{n\to\infty} u_n = \lim_{n\to\infty} (S_{n+1} - S_n) = S - S = 0.$$

Из дальнейших примеров будет видно, что обратное утверждение, вообще говоря, не верно. Однако высказанный критерий полезен, так как дает *необходимое* условие сходимости ряда. Например, то обстоятельство, что ряд, членами которого являются члены геометрической прогрессии со знаменателем $x>1$, расходится, уже вытекает из того, что общий член ряда не стремится к нулю.

Если ряд состоит из положительных членов, то его частичная сумма S_n возрастает вместе с n и может быть только два случая: либо переменная S_n при достаточно большом n делается больше любого наперед заданного числа A, и тогда для последующих значений n она навсегда останется больше A, и $\lim_{n\to\infty} S_n = \infty$, т. е. ряд расходится; либо существует такое число A, что для всех n величина S_n не превышает A, но тогда переменная S_n принуждена будет стремиться к определенному конечному, не превышающему A пределу, и наш ряд будет сходящимся.

Сходимость ряда. Вопрос о том, сходится или расходится данный ряд, часто можно решить, сравнивая его с другим рядом. При этом обычно пользуются следующим критерием.

Если даны два ряда

$$u_0 + u_1 + u_2 + \ldots,$$
$$v_0 + v_1 + v_2 + \ldots$$

с положительными членами, у которых для всех натуральных n, начиная с некоторого значения, имеет место неравенство

$$u_n \leqslant v_n,$$

то сходимость второго ряда влечет за собой сходимость первого, а расходимость первого влечет за собой расходимость второго.

Рассмотрим для примера ряд

$$1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \frac{1}{5} + \frac{1}{6} + \frac{1}{7} + \frac{1}{8} + \ldots,$$

называемый гармоническим. Члены его соответственно не меньше членов ряда

$$1 + \underline{\frac{1}{2}} + \underline{\frac{1}{4} + \frac{1}{4}} + \underline{\frac{1}{8} + \frac{1}{8} + \frac{1}{8} + \frac{1}{8}} + \underbrace{\frac{1}{16} + \ldots + \frac{1}{16}}_{8 \text{ раз}} + \ldots,$$

в котором сумма подчеркнутых каждой чертой слагаемых равна $\frac{1}{2}$.

Очевидно, частичная сумма S_n второго ряда возрастает до бесконечности вместе с n, и, следовательно, гармонический ряд расходится.

Ряд

$$1 + \frac{1}{2^\alpha} + \frac{1}{3^\alpha} + \frac{1}{4^\alpha} + \ldots, \tag{56}$$

где α — положительное число, меньшее единицы, очевидно также расходится, так как для любого n

$$\frac{1}{n^\alpha} > \frac{1}{n} \qquad (0 < \alpha < 1).$$

С другой стороны, можно показать, что ряд (56) при $\alpha > 1$ уже будет сходящимся. Докажем это здесь только при $\alpha \geqslant 2$; для этого рассмотрим ряд

$$\left(1 - \frac{1}{2}\right) + \left(\frac{1}{2} - \frac{1}{3}\right) + \ldots + \left(\frac{1}{n-1} - \frac{1}{n}\right) + \ldots$$

с положительными членами. Он сходится и имеет сумму, равную единице, так как его частичные суммы S_n равны

$$S_n = 1 - \frac{1}{n+1} \to 1 \qquad (n \to \infty).$$

С другой стороны, общий член этого ряда удовлетворяет неравенству

$$\frac{1}{n-1} - \frac{1}{n} = \frac{1}{(n-1)n} > \frac{1}{n^2},$$

откуда следует, что и ряд

$$1+\frac{1}{2^2}+\frac{1}{3^2}+\frac{1}{4^2}+\ldots$$

сходится. Тем более сходится ряд (56) при $\alpha>2$.

Приведем без доказательства еще один часто применяемый признак сходимости и расходимости рядов с положительными членами, называемый признаком Даламбера.

Допустим, что отношение $\frac{u_{n+1}}{u_n}$, когда n стремится к бесконечности, имеет предел q. Тогда при $q<1$ ряд заведомо сходится, при $q>1$ он заведомо расходится, а при $q=1$ вопрос о его сходимости остается открытым.

Обычные суммы конечного числа слагаемых не меняются при их перестановке. Однако это, вообще говоря, не верно для бесконечных рядов. Существуют сходящиеся ряды, в которых возможна перестановка членов, изменяющая их сумму и даже превращающая их в расходящийся ряд. Обладающие такими неустойчивыми суммами ряды теряют одно из основных свойств обычных сумм — свойство перестановочности. Поэтому весьма важно выделить те из рядов, которые сохраняют это свойство. Такими оказываются так называемые абсолютно сходящиеся ряды. Ряд

$$u_0+u_1+u_2+u_3+\ldots$$

называется *абсолютно сходящимся*, если сходится ряд

$$|u_0|+|u_1|+|u_2|+|u_3|+\ldots$$

из абсолютных величин его членов. Можно доказать, что абсолютно сходящийся ряд всегда сходится, т. е. его частичные суммы S_n стремятся к конечному пределу. Всякий сходящийся ряд с членами одного знака, очевидно, абсолютно сходящийся.

Ряд

$$\frac{\sin x}{1^2}+\frac{\sin 2x}{2^2}+\frac{\sin 3x}{3^2}+\ldots$$

может служить примером абсолютно сходящегося ряда, так как ряд

$$\left|\frac{\sin x}{1^2}\right|+\left|\frac{\sin 2x}{2^2}\right|+\left|\frac{\sin 3x}{3^2}\right|+\ldots$$

имеет члены, не превышающие соответственных членов сходящегося ряда

$$1+\frac{1}{2^2}+\frac{1}{3^2}+\ldots$$

Примером сходящегося, но не абсолютно, ряда может служить следующий ряд:

$$1-\frac{1}{2}+\frac{1}{3}-\frac{1}{4}+\ldots$$

Попробуйте доказать это сами.

Ряды функций. Равномерно сходящиеся ряды. В анализе часто приходится иметь дело с рядами, члены которых являются функциями от x:

В предыдущем параграфе у нас уже были примеры таких рядов. Например, ряд $1+x+x^2+x^3+\ldots$ Для одних значений x он сходился, для других — расходился.. Важные в приложениях случаи составляют ряды функций, сходящиеся для всех значений x, принадлежащих к некоторому интервалу, который, в частности, может оказаться всей действительной осью или полуосью. Появляется необходимость такие ряды почленно дифференцировать, интегрировать, выяснять вопрос о непрерывности их суммы и т. д. Если речь идет об обычной сумме конечного числа слагаемых, то у нас есть простые общие правила. Мы знаем, что производная суммы дифференцируемых функций равна сумме их производных, интеграл от суммы непрерывных функций есть сумма их интегралов, сумма непрерывных функций в свою очередь есть непрерывная функция,— все это для конечного числа слагаемых.

Однако при переходе к бесконечным рядам эти простые правила перестают, вообще говоря, выполняться. Можно было бы привести многочисленные примеры сходящихся рядов функций, для которых правила почленного интегрирования и дифференцирования оказываются неверными. Точно так же, когда ряд состоит из непрерывных функций, его сумма может не быть функцией непрерывной. С другой стороны, существует много рядов, которые по отношению к этим правилам ведут себя, как обычные конечные суммы.

Соответствующие глубокие исследования этого вопроса показали, что законность применения этих правил можно заранее гарантировать, если рассматриваемые ряды не только сходятся в каждой отдельной точке рассматриваемого отрезка (промежутка изменения x), но эта сходимость по отношению ко всему промежутку равномерна. Таким образом, в математическом анализе выкристаллизовалось (в середине XIX в.) важное понятие равномерной сходимости ряда.

Рассмотрим ряд

$$S(x)=u_0(x)+u_1(x)+u_2(x)+\ldots,$$

членами которого являются функции, определенные на отрезке $[a, b]$. Будем предполагать, что этот ряд для каждого отдельного значения x из этого отрезка сходится к некоторой сумме $S(x)$. Сумма первых n членов этого ряда

$$S_n(x)=u_0(x)+u_1(x)+\ldots+u_{n-1}(x)$$

есть также некоторая функция от x, определенная на $[a, b]$.

Введем теперь в рассмотрение величину η_n, равную верхней грани значений[1] $|S(x)-S_n(x)|$, когда x изменяется на отрезке $[a, b]$. Эту величину записывают так[2]:

$$\eta_n=\sup_{a\leqslant x\leqslant b}|S_n(x)-S(x)|.$$

[1] См. главу XV (том 3).

[2] sup — сокращение латинского слова superior (наивысший).

В случае, если величина $|S(x) - S_n(x)|$ достигает своего максимального значения, что будет заведомо иметь место, например когда $S(x)$ и $S_n(x)$ непрерывны, то η_n есть просто максимум $|S(x) - S_n(x)|$ на $[a, b]$.

В силу предположенной сходимости нашего ряда для каждого отдельного значения x из отрезка $[a, b]$

$$\lim_{n\to\infty} |S(x) - S_n(x)| = 0.$$

Однако величина η_n может при этом стремиться к нулю, а может и не стремиться. Если величина η_n стремится к нулю при $n \to \infty$, то ряд называется *равномерно сходящимся*. В противном случае ряд называется неравномерно сходящимся. В этом же смысле можно говорить о равномерной и неравномерной сходимости последовательности функций $S_n(x)$, не обязательно связывая их с рядом, частичными суммами которого они являются.

Пример 1. Ряд функций

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

который будем считать определенным только для неотрицательных значений x, т. е. на полупрямой $[0, \infty)$, можно записать в виде

$$\frac{1}{x+1} + \left(\frac{1}{x+2} - \frac{1}{x+1}\right) + \left(\frac{1}{x+3} - \frac{1}{x+2}\right) + \ldots,$$

откуда его частичная сумма равна

$$S_n(x) = \frac{1}{x+n}$$

и

$$\lim_{n\to\infty} S_n(x) = 0.$$

Таким образом, ряд сходится для всех неотрицательных x и имеет сумму $S(x) = 0$. Далее,

$$\eta_n = \sup_{0\leqslant x<\infty} |S_n(x) - S(x)| = \sup_{0\leqslant x<\infty} \frac{1}{x+n} = \frac{1}{n} \to 0 \quad (n \to \infty),$$

и наш ряд равномерно сходится к нулю на полуоси $[0, \infty)$. На рис. 37 изображены графики частичных сумм $S_n(x)$.

Пример 2. Ряд

$$x + x(x-1) + x^2(x-1) + \ldots$$

можно переписать в виде

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

откуда

$$S_n(x) = x^n,$$

следовательно,

$$\lim_{n\to\infty} S_n(x) = \begin{cases} 0, & \text{если } 0 \leqslant x < 1; \\ 1, & \text{если } x = 1. \end{cases}$$

Сумма ряда, таким образом, есть разрывная на отрезке [0, 1] функция с разрывом в точке $x=1$. Величина $|S_n(x)-S(x)|$ при каждом x из [0, 1] меньше единицы, но при x, близких к $x=1$, становится сколь угодно близка к единице. Поэтому

$$\eta_n = \sup_{0\leqslant x\leqslant 1} |S_n(x)-S(x)| = 1$$

для всех $n=1, 2, \ldots$ Таким образом, наш ряд сходится на отрезке [0, 1] неравномерно. На рис. 38 изображены графики функций $S_n(x)$. График суммы ряда состоит из отрезка $0\leqslant x<1$ на оси x-ов без правого конца и из точки (1, 1).

Этот пример показывает, что сумма неравномерно сходящегося ряда непрерывных

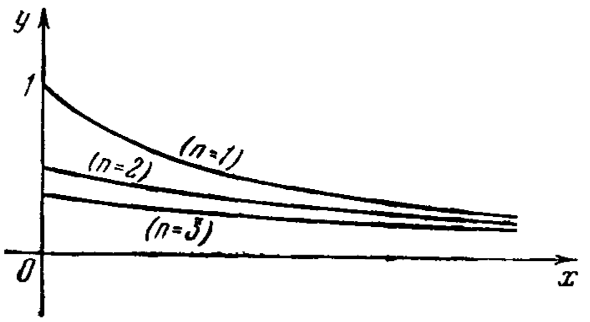


Рис. 37.

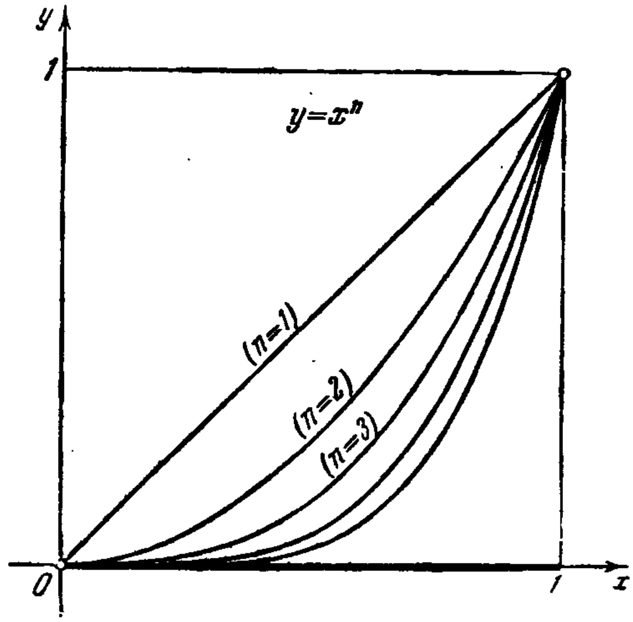


Рис. 38.

функций действительно может оказаться функцией разрывной.

С другой стороны, если наш ряд рассматривать на отрезке $0\leqslant x\leqslant q$, где $q<1$, то

$$\eta_n = \sup_{0\leqslant x\leqslant q} |S_n(x)-S(x)| = \max_{0\leqslant x\leqslant q} x^n = q^n \underset{n\to\infty}{\longrightarrow} 0,$$

и, следовательно, на этом отрезке ряд сходится равномерно, и сумма его, как мы видим, непрерывна. То обстоятельство, что сумма равномерно сходящегося ряда непрерывных функций есть непрерывная функция, как уже говорилось, является общим правилом, которое может быть строго доказано.

Пример 3. Сумма первых n членов ряда $S_n(x)$ имеет график, изображенный жирной ломаной на рис. 39. Очевидно, для всех n будет $S_n(0)=0$; если же $0<x\leqslant 1$, то для всех $n\geqslant\frac{1}{x}$ будем иметь $S_n(x)=0$, и, следовательно, для любого x из отрезка [0, 1]

$$S(x) = \lim_{n\to\infty} S_n(x) = 0.$$

С другой стороны,

$$\eta_n = \sup_{0\leqslant x\leqslant 1} |S_n(x)-S(x)| = \sup|S_n(x)| = n^2.$$

Мы видим, что величина η_n не стремится к нулю; она стремится даже к бесконечности. Заметим, что наш ряд, соответствующий рассматриваемой последовательности $S_n(x)$, нельзя почленно интегрировать на отрезке $[0, 1]$, так как

$$\int_0^1 S(x)\,dx = 0, \qquad \int_0^1 S_n(x)\,dx = \frac{1}{2}n^2\frac{1}{n} = \frac{n}{2},$$

и, таким образом, ряд

$$\int_0^1 S_1(x)\,dx + \int_0^1 [S_2(x) - S_1(x)]\,dx + \int_0^1 [S_3(x) - S_2(x)]\,dx + \ldots$$

сводится к такому расходящемуся ряду:

$$\frac{1}{2} + \left(\frac{2}{2} - \frac{1}{2}\right) + \left(\frac{3}{2} - \frac{2}{2}\right) + \left(\frac{4}{2} - \frac{3}{2}\right) + \ldots$$

Сформулируем без доказательства основные свойства равномерно сходящихся рядов:

1. Сумма равномерно сходящегося на отрезке $[a, b]$ ряда непрерывных функций есть функция, непрерывная на этом отрезке.

2. Если ряд непрерывных функций

$$S(x) = u_0(x) + u_1(x) + u_2(x) + \ldots \tag{57}$$

равномерно сходится на отрезке $[a, b]$, то его можно на этом отрезке почленно интегрировать, т. е. для всяких x_1, x_2 из $[a, b]$ имеет место равенство

$$\int_{x_1}^{x_2} S(t)\,dt = \int_{x_1}^{x_2} u_0(t)\,dt + \int_{x_1}^{x_2} u_1(t)\,dt + \ldots$$

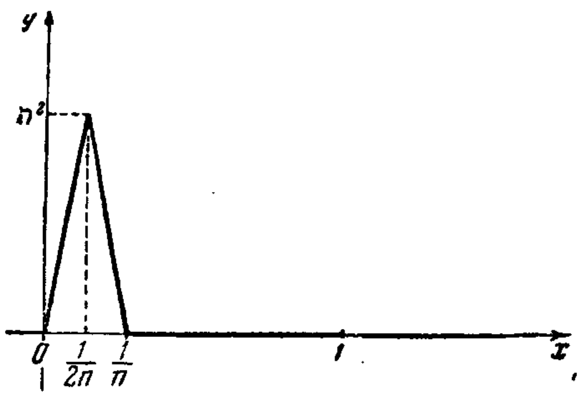


Рис. 39.

3. Пусть на отрезке $[a, b]$ ряд (57) сходится и функции $u_k(x)$ имеют непрерывные производные, тогда равенство

$$S'(x) = u_0'(x) + u_1'(x) + u_2'(x) + \ldots, \tag{58}$$

полученное почленным дифференцированием ряда (57), заведомо будет иметь место на отрезке $[a, b]$, при условии, что получающийся в равенстве (58) справа ряд сходится равномерно.

Степенные ряды. В § 9 мы назвали функцию $f(x)$, заданную на отрезке $[a, b]$, аналитической, если она на этом отрезке имеет производные любого порядка и если в достаточно малой окрестности произвольной точки x_0 отрезка $[a, b]$ функция разлагается в сходящийся к ней ряд Тейлора

$$f(x) = f(x_0) + \frac{f'(x_0)}{1}(x - x_0) + \frac{f''(x_0)}{2!}(x - x_0)^2 + \ldots \tag{59}$$

Если ввести обозначения

$$a_n = \frac{f^{(n)}(x_0)}{n!},$$

то этот ряд может быть записан еще следующим образом

$$f(x) = a_0 + a_1(x - x_0) + a_2(x - x_0)^2 + \ldots \qquad (60)$$

Всякий такой ряд, где числа a_1, a_2, ... — не зависящие от x постоянные, в математике называется *степенным рядом.*

Рассмотрим в качестве примера степенной ряд

$$1 + x + x^2 + x^3 + \ldots, \qquad (61)$$

члены которого образуют геометрическую прогрессию.

Мы знаем, что для всех значений x из интервала $-1 < x < 1$ этот ряд сходится и сумма его равна

$$S(x) = \frac{1}{1-x}.$$

Для остальных значений x ряд расходится.

Легко видеть также, что разность между суммой ряда и суммой его первых n членов выражается формулой

$$S(x) - S_n(x) = \frac{x^n}{1-x}, \qquad (62)$$

и если $-q \leqslant x \leqslant q$, где q — положительное число, меньшее единицы, то

$$\eta_n = \max | S(x) - S_n(x) | = \frac{q^n}{1-q}.$$

Отсюда видно, что η_n стремится к нулю при неограниченном возрастании n и наш ряд равномерно сходится на отрезке $-q \leqslant x \leqslant q$, каково бы ни было положительное число $q < 1$.

Легко проверить, что функция

$$S(x) = \frac{1}{1-x}$$

имеет производную n-го порядка, равную

$$S^{(n)}(x) = \frac{n!}{(1-x)^{n+1}},$$

откуда

$$S^{(n)}(0) = n!$$

и сумма первых n членов формулы Тейлора функции $S(x)$ при $x_0 = 0$ в точности совпадает с суммой первых n членов нашего ряда (59). Мы, кроме того, знаем, что остаточный член формулы, выражаемый равенством (62), стремится к нулю при неограниченном возрастании n для всех x из интервала $-1 < x < 1$. Этим показано, что ряд (61) есть ряд Тейлора своей суммы $S(x)$.

Отметим еще один факт. Выберем на интервале сходимости $-1 < x < 1$ нашего ряда произвольную точку x_0. Легко видеть, что для всех x, достаточно близких к x_0, именно таких, что для них удовлетворяется неравенство

$$\frac{|x-x_0|}{1-x_0}<1,$$

справедливо равенство

$$S(x)=\frac{1}{1-x}=\frac{1}{1-x_0}\frac{1}{\left(1-\frac{x-x_0}{1-x_0}\right)}=\frac{1}{1-x_0}\left[1+\frac{x-x_0}{1-x_0}+\left(\frac{x-x_0}{1-x_0}\right)^2+\ldots\right]=$$

$$=\frac{1}{1-x_0}+\frac{x-x_0}{(1-x_0)^2}+\frac{(x-x_0)^2}{(1-x_0)^3}+\ldots \quad (63)$$

Читатель без особого труда убедится в том, что

$$\frac{S^{(n)}(x_0)}{n!}=\frac{1}{(1-x_0)^{n+1}}.$$

Следовательно, ряд (63) есть ряд Тейлора своей суммы $S(x)$, сходящийся к ней в достаточно малой окрестности произвольной точки x_0, принадлежащей к интервалу сходимости ряда (61), и, значит, ввиду произвольности точки x_0, $S(x)$ есть функция, аналитическая на этом интервале.

Все эти факты, которые мы наблюдали для частного степенного ряда (61), имеют место для произвольных степенных рядов[1]. Именно, каков бы ни был степенной ряд вида (60), где a_k— произвольные, расположенные по любому закону числа, ему соответствует некоторое неотрицательное число R (которое, в частности, может обращаться в ∞), называемое *радиусом сходимости ряда* (60) и обладающее следующими свойствами:

1. Для всех значений x из интервала $x_0-R<x<x_0+R$, который называется интервалом сходимости, ряд сходится, и его сумма $S(x)$ представляет собой на этом интервале аналитическую функцию от x. При этом сходимость равномерна на всяком отрезке $[a, b]$, полностью принадлежащем к интервалу сходимости. Сам ряд представляет собой ряд Тейлора своей суммы.

2. На концах интервала сходимости ряд может сходиться или расходиться, в зависимости от его индивидуальных свойств. Но он заведомо расходится вне замкнутого отрезка $x_0-R\leqslant x\leqslant x_0+R$.

Предлагаем читателю рассмотреть степенные ряды

$$1+\frac{x}{1}+\frac{x^2}{2!}+\frac{x^3}{3!}+\ldots,$$

$$1+x+2!x^2+3!x^3+\ldots,$$

$$1+x+\frac{x^2}{2}+\frac{x^3}{3}+\ldots$$

и убедиться, что первый из них имеет радиус сходимости, равный бесконечности, второй имеет радиус сходимости, равный нулю, а третий — единице.

---

[1] Подробнее об этом см. в главе IX (том 2).

Согласно сделанному выше определению каждая аналитическая функция разлагается в достаточно малой окрестности любой точки, где она задана, в сходящийся к ней степенной ряд. Наоборот, из сказанного следует, что каждый степенной ряд, если его радиус сходимости не равен нулю, имеет в интервале сходимости сумму, представляющую собою аналитическую функцию.

Мы видим, таким образом, что степенные ряды органически связаны с аналитическими функциями. Можно еще сказать, что степенные ряды на интервалах их сходимости являются естественным средством представления аналитических функций, а вместе с этим и естественным средством приближения аналитических функций алгебраическими многочленами[1].

Например, из того, что функция $\frac{1}{1-x}$ разлагается в степенной ряд

$$\frac{1}{1-x}=1+x+x^2+x^3+\ldots,$$

сходящийся на интервале $-1<x<1$, следует, что этот ряд сходится равномерно на любом отрезке $-a\leqslant x\leqslant a$, если $a<1$, а это влечет возможность приближения этой функции на всем отрезке $[-a, a]$ с помощью частичной суммы рассматриваемого ряда с любой наперед заданной точностью.

Допустим, что функцию $\frac{1}{1-x}$ требуется приблизить многочленом на отрезке $\left[-\frac{1}{2}, \frac{1}{2}\right]$ с точностью до 0,01. Замечаем, что для всех x из этого отрезка имеет место неравенство

$$\left|\frac{1}{1-x}-1-x-\ldots-x^n\right|=|x^{n+1}+x^{n+2}+\ldots|\leqslant$$

$$\leqslant|x|^{n+1}+|x|^{n+2}+\ldots\leqslant\frac{1}{2^{n+1}}+\frac{1}{2^{n+2}}+\ldots=\frac{1}{2^n},$$

и так как $2^6=64$, а $2^7=128$, то искомый многочлен, приближающий рассматриваемую функцию на всем протяжении отрезка $\left[-\frac{1}{2}, \frac{1}{2}\right]$ с точностью до 0,01, будет иметь вид

$$\frac{1}{1-x}\approx 1+x+x^2+\ldots+x^7.$$

Отметим еще одно весьма ценное свойство степенных рядов: их всегда можно почленно дифференцировать в интервале сходимости. Этим широко пользуются для решения различных задач в математике.

Пусть например, требуется найти решение дифференциального уравнения $y'=y$ при дополнительном условии $y(0)=1$. Будем искать его решение в виде степенного ряда

$$y=a_0+a_1x+a_2x^2+\ldots$$

[1] Для приближений, выходящих за пределы интервала сходимости степенного ряда, применяются другие методы (см. главу XII, том 2).

В силу дополнительного условия надо считать $a_0 = 1$. Допуская, что этот ряд сходится, мы имеем право его почленно дифференцировать; в результате получим

$$y' = a_1 + 2a_2x + 3a_3x^2 + \dots$$

Подставляя оба эти ряда в дифференциальное уравнение и приравнивая коэффициенты при одинаковых степенях x, получим

$$a_k = \frac{1}{k!} \quad (k = 1, 2, \dots)$$

и искомое решение будет иметь вид

$$y = 1 + \frac{x}{1} + \frac{x^2}{2!} + \frac{x^3}{3!} + \dots$$

Известно, что этот ряд сходится для всех значений x и сумма его равна $y = e^x$.

В данном случае получилось, что сумма ряда есть уже известная нам элементарная функция. Однако это бывает не всегда: может оказаться, что полученный в результате решения задачи сходящийся степенной ряд имеет сумму, не являющуюся элементарной функцией. Таким, например, является ряд

$$y_p(x) = x^p\left[1 - \frac{x^2}{2(2p+2)} + \frac{x^4}{2\cdot 4(2p+2)(2p+4)} - \dots\right],$$

получаемый как решение важного в приложениях дифференциального уравнения Бесселя. Таким образом, степенные ряды служат средством образования новых функций.

ЛИТЕРАТУРА

Популярные брошюры

Болтянский В. Г. Что такое дифференцирование. Гостехиздат, 1955.
Маркушевич А. И. Ряды. Гостехиздат, 1947.
Маркушевич А. И. Площади и логарифмы. Гостехиздат, 1952.
Натансон И. П. Простейшие задачи на максимум и минимум. Гостехиздат, 1950.
Натансон И. П. Суммирование бесконечно малых величин. Гостехиздат, 1953.

Систематические курсы

Бермант А. Ф. Курс математического анализа, ч. I и ч. II. Гостехиздат, 1953.
Смирнов В. И. Курс высшей математики, т. I. Гостехиздат, 1954.
Толстов Г. П. Курс математического анализа, т. I. Гостехиздат, 1954.
Фихтенгольц Г. М. Курс дифференциального и интегрального исчисления, т. I, II, III. Гостехиздат, 1951.
Хинчин А. Я. Краткий курс математического анализа. Гостехиздат, 1953.

«Энциклопедия элементарной математики». Книга третья, функции и пределы. Гостехиздат, 1952.

Первые две статьи содержат учение об элементарных функциях, обстоятельное изложение теории пределов, а также элементарные сведения из дифференциального и интегрального исчисления и теории рядов.

Глава III

АНАЛИТИЧЕСКАЯ ГЕОМЕТРИЯ

§ 1. ВВЕДЕНИЕ

В первой половине XVII в. возникла совершенно новая ветвь математики, так называемая аналитическая геометрия, устанавливающая связь между линиями на плоскости и алгебраическими уравнениями с двумя неизвестными.

Произошел довольно редкий в математике случай: за одно-два десятилетия появилась большая, совсем новая часть математики, основанная притом на очень простой идее, на которую, однако, до того не обращали должного внимания. Появление аналитической геометрии в первой половине XVII в. было не случайным. Переход в Европе к новой, капиталистической форме производства потребовал усовершенствований в целом ряде наук. Только что Галилеем и другими учеными начала создаваться современная механика, во всех областях естествознания накапливались опытные данные, совершенствовались средства наблюдения, вместо устаревших схоластических теорий создавались новые. В астрономии среди передовых ученых окончательно восторжествовало учение Коперника. Мощное развитие дальнего мореплавания настойчиво требовало знания астрономии и начатков механики.

В механике нуждалось военное дело. Эллипсы и параболы, геометрические свойства которых, как конических сечений, были уже подробно известны еще древним грекам почти за 2000 лет перед тем, перестали быть предметами только геометрии, какими они были у греков. После того как Кеплер открыл, что планеты обращаются вокруг Солнца по эллипсам, а Галилей — что брошенный камень летит по параболе, надо было вычислять эти эллипсы, находить те параболы, по которым летят ядра из пушек; надо было отыскать тот закон, по которому убывает с высотой атмосферное давление, открытое Паскалем; надо было фактически вычислять объемы самых различных тел и т. д. и т. п.

Все эти вопросы вызвали к жизни, почти одновременно, развитие трех совсем новых математических наук: аналитической геометрии, дифференциального и интегрального исчислений (включая решение простейших дифференциальных уравнений).

Эти три новые исчисления качественно изменили лицо всей математики. Ее средствам стали доступны задачи, о решении которых до этого и не помышляли.

В первой половине XVII в., т. е. в начале 1600-х годов, ряд наиболее выдающихся математиков был уже близок к идее аналитической геометрии, но было два математика, которые особенно ясно поняли возможность создания новой части математики. Это были Пьер Ферма, советник парламента во французском городе Тулузе, один из крупнейших мировых математиков, и знаменитый французский философ Рене Декарт. Основным создателем аналитической геометрии считают все же Декарта. Только Декарт, как философ, поставил вопрос во всей его общности. Декарт опубликовал большой философский трактат «Рассуждение о методе, чтобы хорошо направлять свой разум и отыскивать истину в науках, с приложениями: диоптрика, метеорология и геометрия».

Последняя часть этого сочинения, озаглавленная «Геометрия», опубликованная в 1637 г., содержит довольно полное, хотя и несколько путаное изложение той математической теории, которую мы с тех пор называем аналитической геометрией.

§ 2. ДВЕ ОСНОВНЫЕ ИДЕИ ДЕКАРТА

Декарт хотел создать метод, который одинаково прилагался бы к решению всех задач геометрии, давал бы, так сказать, общий способ для их решения. Теория Декарта основана на двух идеях: идее координат и идее представлять при помощи координатного метода любое алгебраическое уравнение с двумя неизвестными в виде некоторой линии на плоскости.

Идея координат. *Координатами точки* на плоскости Декарт называет абсциссу и ординату этой точки, т. е. численные величины x и y ее расстояний (с соответственными знаками) до двух выбранных на этой плоскости взаимно перпендикулярных прямых (координатных осей) (см. главу II). Точка пересечения координатных осей, т. е. точка, имеющая координаты (0, 0), называется *началом координат*.

Введением координат Декарт произвел, как говорят, «арифметизацию» плоскости. Вместо того, чтобы как-либо геометрически указывать точку, достаточно задать пару чисел x, y, и обратно (рис. 1).

Идея сопоставления уравнениям с двумя неизвестными линий на плоскости. Вторая идея Декарта следующая. До Декарта, в случае если имелось одно алгебраическое уравнение с двумя неизвестными $F(x, y) = 0$, говорили, что задача неопределенная, так как неизвестных этих нельзя из него определить хотя бы потому, что за одно из них, например за x, можно взять любое число и, подставив это число вместо x, получить уже одно уравнение с одним неизвестным y, которое, вообще говоря, из него и можно найти. И тогда это *произвольно взятое* x вместе с *так полученным* y

будет удовлетворять заданному уравнению. Поэтому такое «неопределенное» уравнение не считали интересным.

Декарт посмотрел на дело иначе. Он предложил в уравнении с двумя неизвестными x считать абсциссой точки, а соответственное ему y — ее ординатой. Тогда, если менять непрерывно x, то для каждого из этих x вычисляется из уравнения свое вполне определенное y, и, следовательно, получается, вообще говоря, совокупность точек, образующих линию (рис. 2) [1].

Таким образом, всякому алгебраическому уравнению с двумя переменными $F(x, y) = 0$ соответствует некоторая вполне определенная линия на плоскости, а именно линия, представляющая собою совокупность

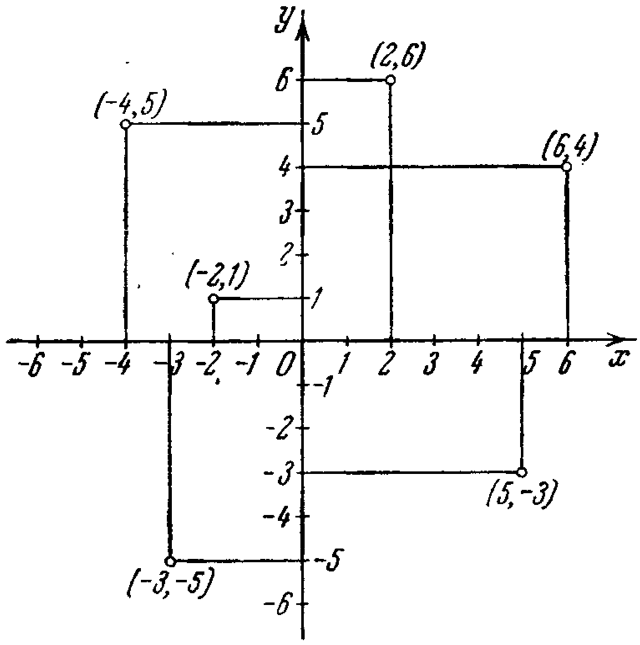


Рис. 1.

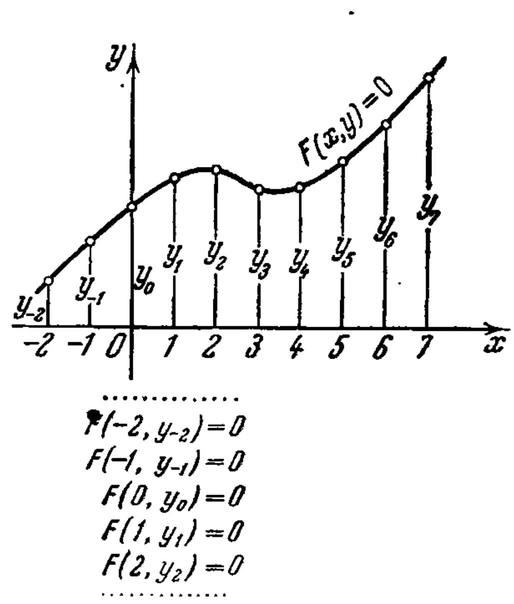


Рис. 2.

всех тех точек плоскости, координаты которых удовлетворяют уравнению $F(x, y) = 0$.

Это замечание Декарта *открывало целую новую науку.*

Основные задачи, решаемые аналитической геометрией, и определение аналитической геометрии. Аналитическая геометрия дает возможность: 1) задачи на построение решать вычислением (см., например, деление отрезка в данном отношении, стр. 183); 2) находить уравнения кривых, заданных каким-либо геометрическим свойством (например, из условия постоянства суммы расстояний до двух точек получить уравнение эллипса, стр. 191); 3) доказывать новые геометрические теоремы алгебраически (см., например, вывод теории диаметров Ньютона, стр. 189); 4) наоборот, представляя алгебраическое уравнение геометрически, выяснять его

[1] Иногда уравнению не удовлетворяет никакая точка (x, y) с действительными координатами, иногда всего одна или несколько таких точек. В таком случае говорят, что линия мнимая или вырождается в точки (см. стр. 202).

алгебраические свойства (см., например, решение уравнения 3-й и 4-й степени при помощи пересечения параболы с окружностью, стр. 186—187).

Аналитическая геометрия есть, таким образом, та часть математики, которая, применяя координатный метод, исследует геометрические объекты средствами алгебры.

§ 3. ПРОСТЕЙШИЕ ЗАДАЧИ

Координаты точки, делящей отрезок в данном отношении. Зная координаты (x_1, y_1) и (x_2, y_2) двух точек M_1 и M_2, найти координаты (x, y) точки M, делящей отрезок M_1M_2 в отношении m к n (рис. 3). Из подобия заштрихованных треугольников получаем:

$$\frac{x-x_1}{x_2-x}=\frac{m}{n}, \text{ откуда } x=\frac{nx_1+mx_2}{m+n},$$

$$\frac{y-y_1}{y_2-y}=\frac{m}{n}, \text{ откуда } y=\frac{ny_1+my_2}{m+n}.$$

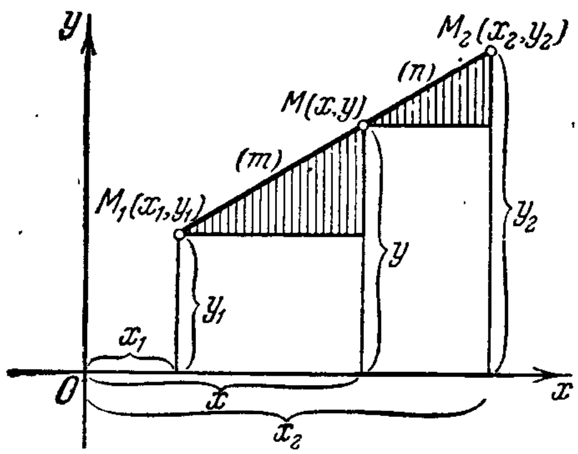


Рис. 3.

Расстояние между двумя точками. Найти расстояние d между точками M_1 и M_2, координаты которых (x_1, y_1) и (x_2, y_2). Из заштрихованного прямоугольного треугольника (рис. 4) получается по теореме Пифагора:

$$d=\sqrt{(x_2-x_1)^2+(y_2-y_1)^2}.$$

Площадь треугольника. Найти площадь S треугольника $M_1M_2M_3$ (рис. 5), если координаты его вершин соответственно (x_1, y_1), (x_2, y_2),

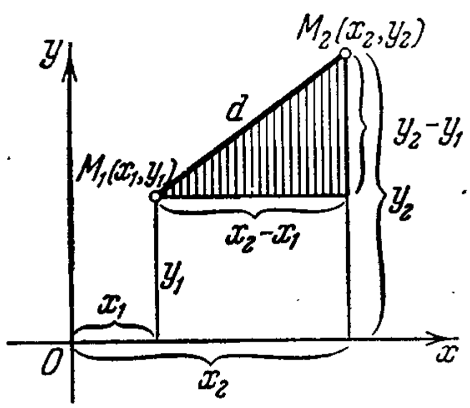


Рис. 4.

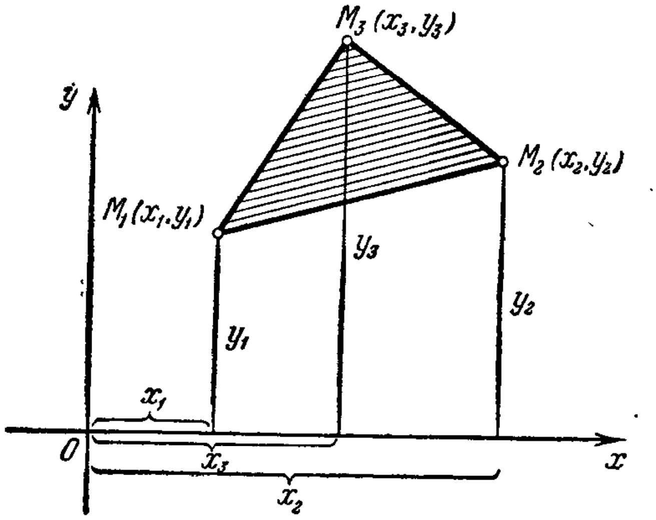


Рис. 5.

(x_3, y_3). Рассматривая треугольник как сумму площадей трапеций с основаниями y_1, y_3, и y_3, y_2 минус трапеция с основаниями y_1, y_2 и переписывая произведение — $(y_1 + y_2)(x_2 - x_1)$ в виде $(y_1 + y_2)(x_1 - x_2)$, получаем:

$$S = \frac{1}{2}[(y_1 + y_2)(x_1 - x_2) + (y_2 + y_3)(x_2 - x_3) + (y_3 + y_1)(x_3 - x_1)].$$

В этих задачах надо только еще проверить, что полученные формулы без всяких изменений остаются верными и для тех случаев, когда одна или несколько координат или их разностей отрицательны. Такую проверку нетрудно сделать.

Отыскание точек пересечения двух линий. Опираясь на вторую основную идею, что уравнение $F(x, y) = 0$ выражает линию, особенно просто находить точки пересечения двух линий. Чтобы найти координаты точек пересечения двух линий, надо, очевидно, совместно решить уравнения, их выражающие. Получившаяся из обычного решения этих двух уравнений пара чисел x, y будет давать точку, координаты которой удовлетворяют как одному, так и другому уравнению, т. е. точку, которая лежит как на первой, так и на второй линии, — это и есть точка их пересечения.

Решение геометрических задач средствами аналитической геометрии, как видим, очень удобно для практики, в частности потому, что все решения сразу получаются в готовом виде в *числах. Такой геометрии, такой науки как раз и не хватало той эпохе.*

§ 4. ИССЛЕДОВАНИЕ ЛИНИЙ, ВЫРАЖЕННЫХ УРАВНЕНИЯМИ 1-Й И 2-Й СТЕПЕНИ

Уравнение 1-й степени. При использовании второй идеи Декарт прежде всего рассмотрел, какие линии соответствуют уравнению 1-й степени, т. е. уравнению

$$Ax + By + C = 0, \tag{1}$$

где A, B, C — некоторые численные коэффициенты, такие что A и B не равны одновременно нулю. Оказалось, что на плоскости такому уравнению всегда соответствует прямая линия.

Докажем, что уравнение (1) всегда выражает прямую, и обратно, что всякой прямой на плоскости соответствует вполне определенное уравнение (1). Действительно, пусть например, $B \neq 0$, тогда уравнение (1) можно разрешить относительно y

$$y = kx + l,$$

где $k = -\frac{A}{B}$; $l = -\frac{C}{B}$.

Рассмотрим сначала уравнение $y = kx$. Оно выражает, очевидно, прямую, проходящую через начало координат и составляющую угол φ с осью x, тангенс которого $\operatorname{tg}\varphi$ есть k (рис. 6). Действительно, это уравнение можно переписать так: $\frac{y}{x} = k$, и все точки (x, y) такой прямой удовлетворяют этому уравнению, а никакая точка $(x, \bar{y})$, не лежащая на этой прямой, уже этому уравнению не удовлетворяет, так как для нее будет $\frac{\bar{y}}{x}$ либо больше, либо меньше k. При этом, если $\operatorname{tg}\varphi > 0$, то для такой прямой x и y оба положительны или оба отрицательны, а если $\operatorname{tg}\varphi < 0$, то знаки их противоположны.

Итак, уравнение $y = kx$ выражает прямую, проходящую через начало координат O, а, следовательно, уравнение $y = kx + l$ выражает тоже прямую, а именно ту, которая получается из предыдущей, если ее перенести параллельно самой себе так, чтобы ординаты всех ее точек увеличились на l (рис. 7).

Рассмотренные выше формулы: координат точки, делящей отрезок в данном отношении, расстояния между двумя заданными точками,

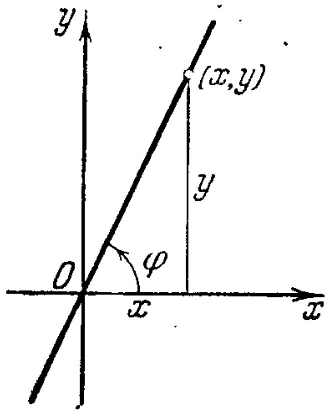


Рис. 6.

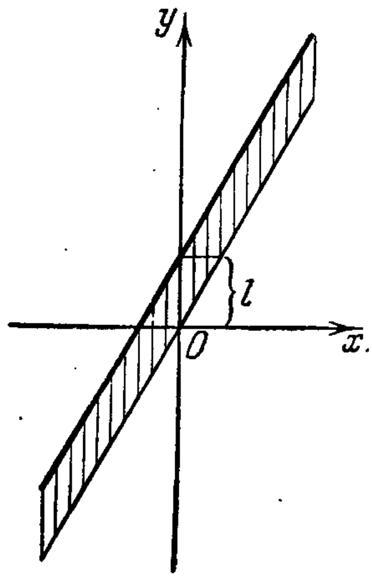


Рис. 7.

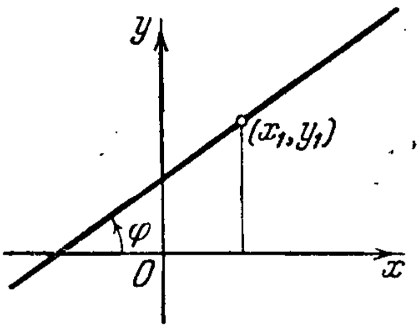


Рис. 8.

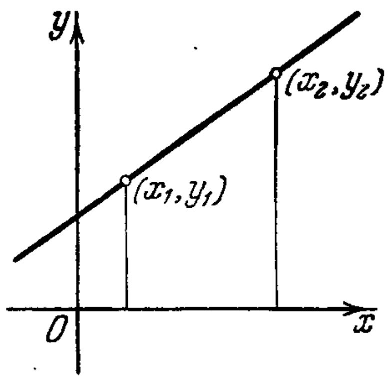


Рис. 9.

площади треугольника, а также сведения об уравнении прямой дают уже возможность решать множество задач.

Уравнение прямой, проходящей через одну или две заданные точки. Пусть M_1 — некоторая точка, имеющая координаты x_1 и y_1, и k — некоторое заданное число. Уравнение $y = kx + l$ выражает прямую, образующую с осью Ox угол, тангенс которого равен k, и отсекающую на оси Oy отрезок, равный l. Подберем l так, чтобы эта прямая проходила через точку (x_1, y_1). Для этого координаты точки M_1 должны удовлетворять уравнению, т. е. должно быть $y_1 = kx_1 + l$, откуда $l = y_1 - kx_1$.

Подставляя это l, мы получаем уравнение прямой, проходящей через заданную точку (x_1, y_1) и образующей с осью Ox угол, тангенс которого равен k (рис. 8). Оно будет $y = kx + y_1 - kx_1$ или

$$y - y_1 = k(x - x_1).$$

П р и м е р. Пусть угол между прямой и осью Ox равен 45°, а точка M имеет координаты (3, 7), тогда уравнение соответствующей прямой (так как $\operatorname{tg} 45° = 1$) будет: $y - 7 = 1(x - 3)$ или $x - y + 4 = 0$.

Если потребовать, чтобы прямая, проходящая через точку (x_1, y_1), проходила еще и через точку (x_2, y_2), то получим условие $y_2 - y_1 = k(x_2 - x_1)$, налагаемое на k. Найдя отсюда k и подставив его в предыдущее уравнение, получим уравнение прямой, проходящей через две заданные точки (рис. 9)

$$\frac{x - x_1}{x_2 - x_1} = \frac{y - y_1}{y_2 - y_1}.$$

Результат Декарта относительно уравнения 2-й степени. Декарт исследовал также вопрос о том, какие линии на плоскости выражает уравнение 2-й степени с двумя переменными, общий вид которого

$$Ax^2 + Bxy + Cy^2 + Dx + Ey + F = 0,$$

и показал, что такое уравнение, вообще говоря, выражает эллипсы, гиперболы или параболы — кривые, хорошо известные математикам древности.

Таковы важнейшие достижения Декарта. Этим, однако, книга Декарта далеко не ограничивалась; Декарт еще исследует уравнения ряда интересных геометрических мест, рассматривает теоремы о преобразовании алгебраических уравнений, приводит без доказательства свое знаменитое *правило знаков* для отыскания числа положительных корней уравнения, все корни которого действительные (см. главу IV, § 4), и, наконец, предлагает замечательный способ для отыскания действительных корней уравнений 3-й и 4-й степени при помощи пересечения параболы $y = x^2$ с окружностями.

§ 5. МЕТОД ДЕКАРТА ДЛЯ РЕШЕНИЯ АЛГЕБРАИЧЕСКИХ УРАВНЕНИЙ 3-й и 4-й СТЕПЕНИ

Преобразование уравнений 3-й и 4-й степени к уравнению 4-й степени, не имеющему члена с x^3. Покажем, что решение любого уравнения 3-й и 4-й степени можно свести к решению уравнения вида

$$x^4 + px^2 + qx + r = 0. \tag{2}$$

Пусть задано уравнение 3-й степени $z^3 + az^2 + bz + c = 0$. Положив $z = x - \frac{a}{3}$, мы получим $\left(x - \frac{a}{3}\right)^3 + a\left(x - \frac{a}{3}\right)^2 + b\left(x - \frac{a}{3}\right) + c = 0$. Члены с x^2 при раскрытии скобок взаимно уничтожаются, и мы получим уравнение $x^3 + px + q = 0$. Помножив это уравнение на x, что добавит еще один корень $x_4 = 0$, мы приведем его к виду (2), где $r = 0$.

Уравнение 4-й степени $z^4+az^3+bz^2+cz+d=0$, можно привести к виду (2), положив $z=x-\frac{a}{4}$. Решение всякого уравнения 3-й и 4-й степени можно, следовательно, свести к решению уравнения вида (2).

Решение уравнений 3-й и 4-й степени пересечением окружности с параболой $y=x^2$. Составим уравнение окружности с центром (a, b) и радиусом R. Если (x, y) — некоторая точка, то квадрат ее расстояния до точки (a, b) равен $(x-a)^2+(y-b)^2$ (см. § 3). Уравнение рассматриваемой окружности, следовательно, есть

$$(x-a)^2+(y-b)^2=R^2.$$

Попробуем теперь найти точки пересечения этой окружности с параболой $y=x^2$. Для этого, в силу сказанного в § 3, надо совместно решить уравнение этой окружности и уравнение параболы:

$$x^2+y^2-2ax-2by+a^2+b^2-R^2=0,$$
$$y=x^2.$$

Подставляя y из второго уравнения в первое, мы получим относительно x уравнение 4-й степени

$$x^2+x^4-2ax-2bx^2+a^2+b^2-R^2=0$$

или уравнение

$$x^4+(1-2b)x^2-2ax+a^2+b^2-R^2=0.$$

Если взять a, b и R^2 такие, чтобы было

$$1-2b=p,\quad -2a=q,\quad a^2+b^2-R^2=r,$$

то получится как раз уравнение (2). Для этого надо взять

$$a=-\frac{q}{2},\quad b=\frac{1-p}{2},\quad R^2=\frac{q^2}{4}+\frac{(1-p)^2}{4}-r. \tag{3}$$

В последней формуле (3), вообще говоря, R^2 может оказаться отрицательным. Однако в случае, когда уравнение (2) имеет хотя бы один действительный корень x_1, имеет место равенство

$$x_1^4+(1-2b)x_1^2-2ax_1+a^2+b^2-R^2=0. \tag{4}$$

Обозначая x_1^2 через y_1, равенство (4) можно переписать так:

$$x_1^2+y_1^2-2ax_1-2by_1+a^2+b^2-R^2=0$$

или так:

$$(x_1-a)^2+(y_1-b)^2=R^2.$$

Таким образом, в случае, когда уравнение (2) имеет действительный корень, число $R^2=\frac{(1-p)^2+q^2}{4}-r$ *положительное*, уравнение

$$(x-a)^2+(y-b)^2=R^2$$

есть уравнение окружности, и все действительные корни уравнения (2) суть абсциссы точек пересечения параболы $y=x^2$ с этой окружностью.

(В случае $r = 0$, $R^2 = a^2 + b^2$, эта окружность проходит через начало координат).

Таким образом, если заданы коэффициенты p, q, r уравнения (2) и надо найти по формулам (3) a, b и R^2, то, если $R^2 < 0$, уравнение (2) заведомо не имеет действительных корней. Если же $R^2 \geqslant 0$, то абсциссы точек пересечения окружности с центром (a, b) и радиусом R с раз навсегда вычерченной параболой $y = x^2$ дают все действительные корни уравнения (2), причем и в случае $R^2 > 0$ получаемая окружность может не пересекать параболы и уравнение (2) может не иметь действительных корней.

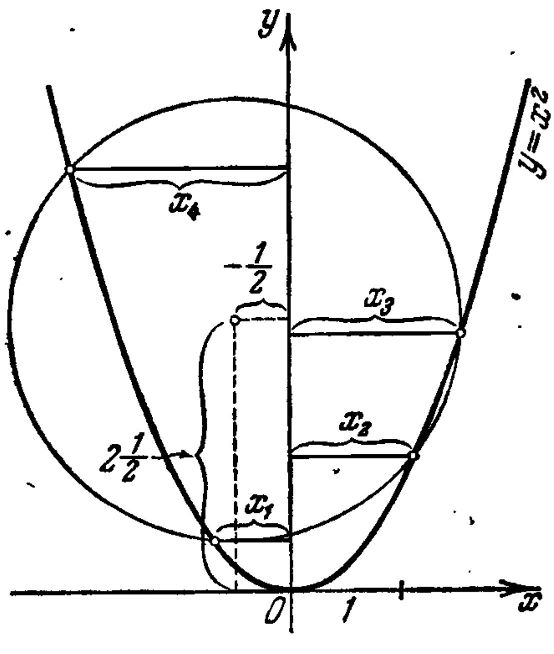


Рис. 10.

Пример. Пусть задано уравнение 4-й степени:

$$x^4 - 4x^2 + x + \frac{5}{2} = 0.$$

Мы имеем тогда

$$a = -\frac{1}{2}, \; b = \frac{5}{2} = 2\frac{1}{2},$$

$$R = \sqrt{\frac{1}{4} + \frac{25}{4} - \frac{5}{2}} = 2.$$

На рис. 10 показаны соответствующая окружность и корни x_1, x_2, x_3, x_4 рассматриваемого уравнения.

Параграфы 1, 2, 3 и 4 содержат в кратком и более близком к современному изложении наиболее существенное, что есть в книге Декарта.

Со времени Декарта до наших дней аналитическая геометрия прошла большой путь развития, плодотворный для самых разных частей математики. Мы попробуем в дальнейших параграфах этой главы проследить наиболее важные этапы этого пути.

Прежде всего надо сказать, что изобретатели анализа бесконечно малых уже владели методом Декарта. Будь то вопрос о касательных или нормалях (перпендикулярах к касательным в точках касания) к линии, или вопрос о максимумах и минимумах функции, если рассматривать его геометрически, или вопрос о радиусе кривизны линии в данной ее точке и т. д., — всегда рассматривают прежде всего, по Декарту, уравнение этой линии и затем уже находят уравнение нормали, уравнение касательной и т. д. Поэтому анализ бесконечно малых, дифференциальное и интегральное исчисления были бы немыслимы без предварительной разработки аналитической геометрии.

§ 6. ОБЩАЯ ТЕОРИЯ ДИАМЕТРОВ НЬЮТОНА

Первый, кто в самой аналитической геометрии сделал дальнейший шаг вперед, был Ньютон. В 1704 г. он рассмотрел теорию линий 3-го порядка, т. е. линий, выражаемых алгебраическими уравнениями 3-й степени с двумя неизвестными. В той же работе Ньютон нашел, между прочим, изящную общую теорему о «диаметрах», соответствующих секущим данного направления. Он показал следующее.

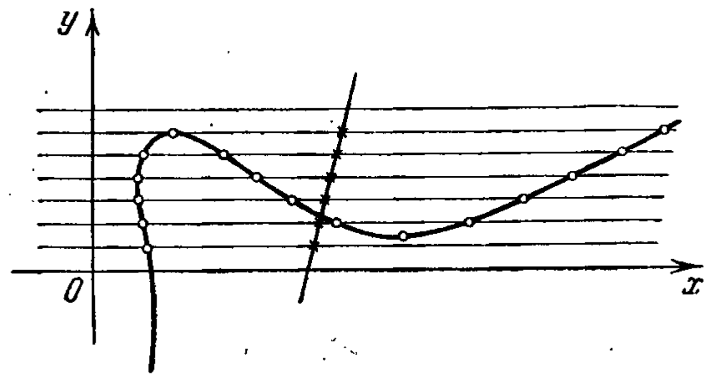


Рис. 11.

Пусть дана линия n-го порядка, т. е. линия, выражаемая алгебраическим уравнением с двумя переменными n-й степени, тогда любая прямая, секущая ее, имеет, вообще говоря, с ней n общих точек. Пусть M — точка секущей, которая является «центром тяжести» этих точек ее пересечения с рассматриваемой линией n-го порядка, т. е. центром тяжести совокупности n равных друг другу точечных масс, расположенных в этих точках. Оказывается, если брать всевозможные параллельные друг другу секущие и для каждой из них рассматривать этот центр тяжести M, то все эти точки M лежат на одной прямой. Эту прямую Ньютон назвал «диаметром» линии n-го порядка, соответствующим данному наклону секущих. Доказательство этой теоремы при помощи аналитической геометрии вовсе нетрудно, поэтому мы его приведем.

Пусть дана линия n-го порядка и некоторые параллельные друг другу ее секущие. Возьмем такие оси координат, чтобы эти секущие были параллельны оси Ox (рис. 11). Тогда их уравнения будут иметь вид $y = l$, где l для разных этих секущих — разные постоянные. Пусть $F(x, y) = 0$ то уравнение, которое выражает рассматриваемую линию n-го порядка в этих координатных осях. Легко показать, что при переходе от одних прямоугольных осей к другим, хотя уравнение линии и меняется, но порядок его не меняется (это будет сделано в § 8). Поэтому $F(x, y)$ будет тоже многочлен n-й степени. Для нахождения абсцисс точек пересечения нашей линии с секущей $y = l$ надо совместно решить уравнения $F(x, y) = 0$ и $y = l$, в результате чего получится уравнение, вообще говоря, n-й степени относительно x

$$F(x, l) = 0, \tag{5}$$

из которого находят абсциссы $x_1, x_2, \ldots, x_n$. Абсцисса x_c центра тяжести n точек пересечения по самому определению центра тяжести равна

$$x_c = \frac{x_1 + x_2 + \ldots + x_n}{n}.$$

Но, как известно из теории алгебраических уравнений, сумма $x_1 + x_2 + \ldots + x_n$ корней уравнения равна коэффициенту при $(n-1)$-й степени неизвестной x, взятому с обратным знаком, разделенному на коэффициент при n-й степени. А так как сумма показателей над x и y в каждом члене $F(x, y)$ равна n или меньше, то член с x^n вовсе не содержит y и имеет вид Ax^n, где A — постоянное, а члены с x^{n-1} если и содержат y, то

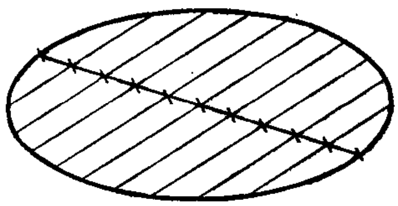


Рис. 12.

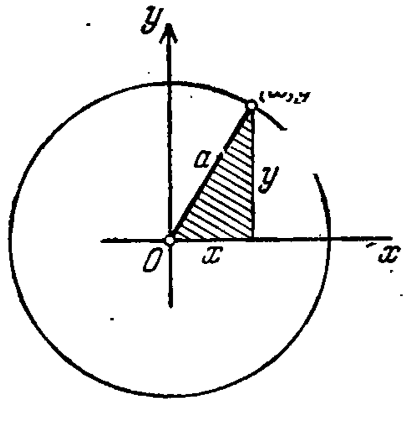


Рис. 13.

не выше чем в 1-й степени, т. е. имеют вид $x^{n-1}(By + C)$. Следовательно, коэффициент при x^n есть A, а при x^{n-1} есть $Bl + C$, и мы имеем для данного l

$$x_c = -\frac{Bl + C}{nA}.$$

Но секущая параллельна оси Ox, и во всех ее точках $y = l$, а следовательно, и ордината y центра тяжести точек ее пересечения с рассматриваемой линией n-го порядка также равна l; таким образом мы получим окончательно $nAx_c + By_c + C = 0$, т. е. координаты x_c, y_c всех рассматриваемых центров тяжести всех этих секущих удовлетворяют уравнению 1-й степени, и, следовательно, эти центры тяжести лежат на прямой.

Случай, когда в $F(x, y)$ не входит x^n, исследуется аналогично.

В случае линий 2-го порядка ($n = 2$) центр тяжести двух точек есть просто середина между ними, и получается, что геометрическое место середин параллельных хорд линии 2-го порядка есть прямая (рис. 12), что было как для эллипса, так для гиперболы и параболы уже известно древним. Но ими это доказывалось, даже и для этих частных случаев, довольно трудными геометрическими рассуждениями, а тут новая, не известная древним, общая теорема доказана совсем просто.

Такие примеры обнаруживают силу аналитической геометрии.

§ 7. ЭЛЛИПС, ГИПЕРБОЛА И ПАРАБОЛА

В этом и следующем параграфах мы рассмотрим линии 2-го порядка. Прежде чем исследовать общее уравнение 2-й степени, полезно рассмотреть некоторые простейшие его виды.

Уравнение окружности с центром в начале координат. Рассмотрим прежде всего уравнение

$$x^2 + y^2 = a^2.$$

Оно, очевидно, выражает окружность с центром в начале координат и радиусом a, что следует (рис. 13) из теоремы Пифагора для заштрихованного прямоугольного треугольника, так как, какую бы точку (x, y) этой окружности ни взять, ее x и y удовлетворяют этому уравнению, и, наоборот, если координаты x, y точки удовлетворяют этому уравнению, то она принадлежит этой окружности, т. е. эта окружность есть совокупность всех точек плоскости, удовлетворяющих этому уравнению.

Уравнение эллипса и его фокальное свойство. Пусть заданы две точки F_1 и F_2, расстояние между которыми равно $2c$. Найдем уравнение геометрического места всех точек M плоскости, сумма расстояний от которых до точек F_1 и F_2 равна некоторой постоянной величине $2a$ (где a, конечно, больше c). Такая линия называется *эллипсом*, а точки F_1 и F_2 его *фокусами*.

Выберем прямоугольную систему координат так, чтобы точки F_1 и F_2 лежали на оси Ox и начало координат лежало посредине между ними. Тогда координаты точек F_1 и F_2 будут $(c, 0)$ и $(-c, 0)$. Возьмем произвольную точку M с координатами (x, y), принадлежащую исследуемому геометрическому месту, и напишем, что сумма расстояний от нее до точек F_1 и F_2 равна $2a$

$$\sqrt{(x-c)^2 + (y-0)^2} + \sqrt{(x+c)^2 + (y-0)^2} = 2a. \tag{6}$$

Этому уравнению удовлетворяют координаты (x, y) любой точки рассматриваемого геометрического места. Очевидно и обратное, а именно, что любая точка, координаты которой удовлетворяют уравнению (6) принадлежит этому геометрическому месту. Уравнение (6) есть, таким образом, уравнение рассматриваемого геометрического места. Остается его упростить.

Возведя обе части в квадрат, получим

$$x^2 - 2cx + c^2 + y^2 + 2\sqrt{(x^2 - 2cx + c^2 + y^2)(x^2 + 2cx + c^2 + y^2)} + \\ + x^2 + 2cx + c^2 + y^2 = 4a^2,$$

или после упрощения

$$x^2 + y^2 + c^2 - 2a^2 = -\sqrt{(x^2 + y^2 + c^2)^2 - 4c^2x^2}.$$

Возведя снова обе части в квадрат, получим

$$(x^2 + y^2 + c^2)^2 - 4a^2(x^2 + y^2 + c^2) + 4a^4 = (x^2 + y^2 + c^2)^2 - 4c^2x^2$$

или после упрощения

$$(a^2 - c^2)x^2 + a^2y^2 = (a^2 - c^2)a^2.$$

Положив здесь $a^2 - c^2 = b^2$ (что можно сделать, так как $a > c$), получаем $b^2x^2 + a^2y^2 = a^2b^2$, а разделив всё на a^2b^2, будем иметь

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1 . \tag{7}$$

Координаты (x, y) любой точки M рассматриваемого геометрического места удовлетворяют, таким образом, уравнению (7).

Можно показать, что и наоборот, если координаты некоторой точки удовлетворяют уравнению (7), то они удовлетворяют и уравнению (6).

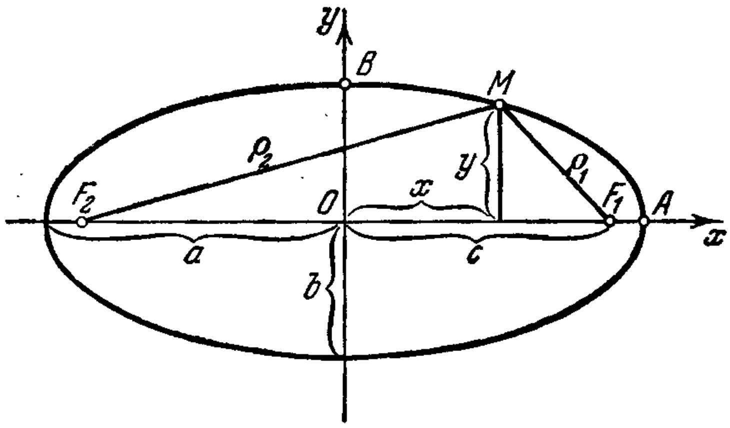


Рис. 14.

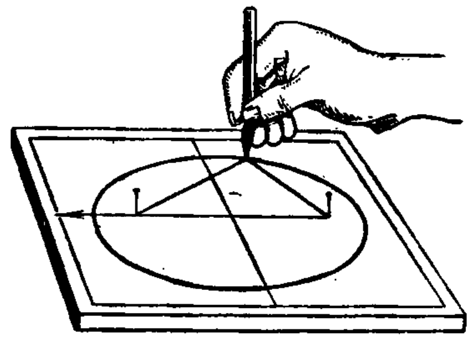


Рис. 15.

Уравнение (7) есть, следовательно, уравнение этого геометрического места, т. е. уравнение эллипса (рис. 14).

Проведенные рассуждения являются классическим примером отыскания уравнения линии, заданной некоторым ее геометрическим свойством.

На свойстве эллипса, что сумма расстояний любой его точки от двух данных точек есть величина постоянная, основан известный способ чертить эллипс при помощи нитки (рис. 15).

З а м е ч а н и е. Для определения того, какая линия называется эллипсом, можно было бы взять не рассмотренное здесь фокальное свойство эллипса, а какое-либо другое геометрическое свойство, его характеризующее, например то, что эллипс есть результат «равномерного сжатия» окружности к ее диаметру (см. стр. 220) или какое-либо иное.

Положив в уравнении (7) эллипса $y = 0$, мы получим $x = \pm a$, т. е. a есть длина отрезка OA (см. рис. 14), называемого *большой полуосью* эллипса. Аналогично положив $x = 0$, получим, что $y = \pm b$, т. е., что b — длина отрезка OB, называемого *малой полуосью* эллипса.

Число $\varepsilon = \frac{c}{a}$ называется *эксцентриситетом* эллипса, причем так как $c = \sqrt{a^2 - b^2} < a$, то эксцентриситет эллипса меньше единицы. В случае окружности $c = 0$, и следовательно $\varepsilon = 0$; фокусы оба в одной точке — в центре окружности (так как $OF_1 = OF_2 = 0$), но предыдущий способ черчения ниткой все же сохраняется.

Законы движения планет. Изучая обширный материал наблюдений Тихо Браге над движением планеты Марс по небесной сфере, Кеплер обнаружил, что планеты обращаются вокруг Солнца по эллипсам, так что Солнце находится в одном из фокусов этого эллипса (другой фокус ничем не занят и в движении планеты вокруг Солнца никакой роли не играет) (рис. 16), и что при этом фокальный радиус ρ в одинаковые времена зачерчивает секторы одинаковой площади [1], а Ньютон показал, что необходимость такого движения математически следует из закона инерции, закона пропорциональности ускорения действующей силе и закона всемирного тяготения.

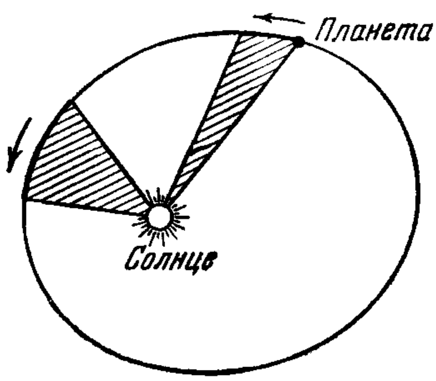


Рис. 16.

Эллипс инерции. В качестве примера на применение эллипса в технической задаче рассмотрим так называемый эллипс инерции пластинки.

Пусть имеется некоторая пластинка равномерной толщины и из однородного материала, например цинковая пластинка какой-нибудь формы. Будем вращать ее вокруг какой-нибудь оси, лежащей в ее плоскости. Тело, двигающееся

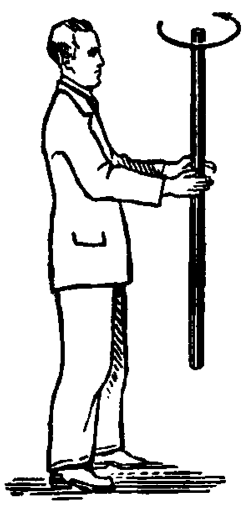


Рис. 17а.

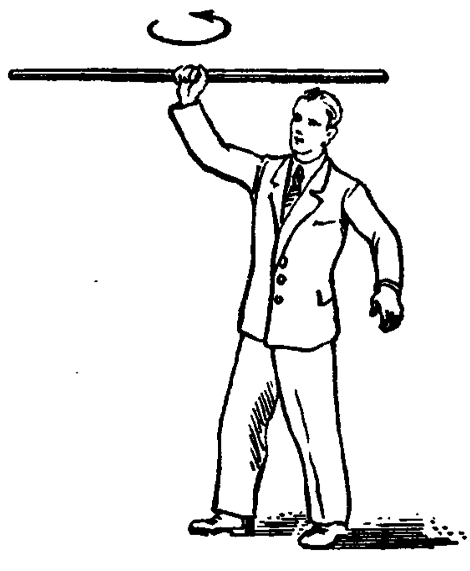


Рис. 17б.

прямолинейно, имеет, как известно, некоторую инерцию по отношению к этому прямолинейному движению, пропорциональную его массе (независимо от формы тела и расположения его масс). Подобно этому и тело, вращающееся вокруг оси, например маховик, имеет инерцию по отношению к этому вращению. Однако в случае вращения инерция пропорциональна не только массе вращаемого тела, но зависит также от расположения масс этого тела по отношению к оси вращения, так как

[1] Эксцентриситеты планетных орбит очень невелики, так что орбиты плаиет почти окружности.

инерция по отношению к вращению больше, если массы более удалены от оси. Например, совсем легко привести сразу в быстрое вращение палку вокруг ее продольной оси (рис. 17а). Если же попытаться сразу привести ее в быстрое вращение вокруг оси, перпендикулярной к ее длине, хотя бы и проходящей через ее середину, придется затратить, если эта палка не очень легкая, значительное усилие (рис. 17б).

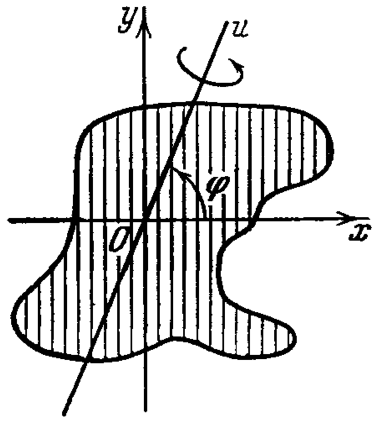


Рис. 18.

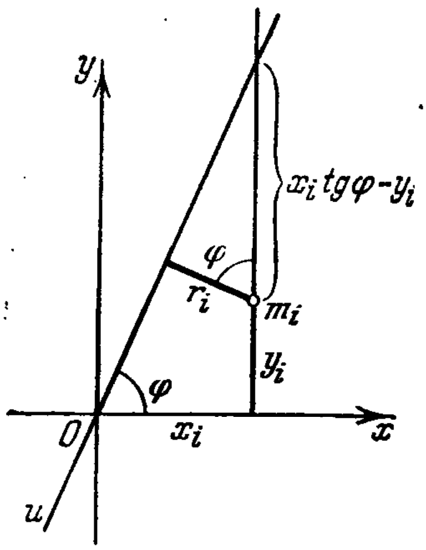


Рис. 19.

Можно показать, что инерция по отношению к вращению тела вокруг некоторой оси, так называемый «момент инерции» тела относительно этой оси, равен $\sum r_i^2 m_i$. (Если под $\sum r_i^2 m_i$ понимать сокращенную запись суммы $r_1^2 m_1 + r_2^2 m_2 + \ldots + r_n^2 m_n$ и мыслить тело разбитым на весьма малые по своей величине элементы, причем m_i — масса элемента, а r_i — расстояние элемента от оси вращения и суммирование распространено на все элементы тела.)

Вернемся к нашей пластинке. Пусть O (рис. 18) — какая-нибудь точка этой пластинки. Рассмотрим моменты инерции J_u этой пластинки по отношению к различным осям u, проходящим через точку O и лежащим в плоскости пластинки. Для этого, приняв точку O за начало прямоугольных координат, выберем любым образом оси Ox и Oy в плоскости пластинки и будем ось вращения u характеризовать углом ее φ с осью Ox. Легко видеть (рис. 19), что

$$r_i = |(x_i \operatorname{tg}\varphi - y_i)\cos\varphi| = |x_i \sin\varphi - y_i \cos\varphi|.$$

Отсюда

$$\sum r_i^2 m_i = \sum (x_i^2 \sin^2\varphi - 2x_i y_i \sin\varphi\cos\varphi + y_i^2\cos^2\varphi)\, m_i =$$
$$= \sin^2\varphi \sum x_i^2 m_i - 2\sin\varphi\cos\varphi \sum x_i y_i m_i + \cos^2\varphi \sum y_i^2 m_i.$$

Величины $\sin^2\varphi$, $2\sin\varphi\cos\varphi$ и $\cos^2\varphi$ вынесены за знаки сумм, так как это — величины, постоянные для данной оси u. Обозначим теперь

$$\sum x_i^2 m_i = A, \qquad -\sum x_i y_i m_i = B, \qquad \sum y_i^2 m_i = C.$$

Величины A, B и C не зависят от выбора оси u, а зависят только от формы пластинки и расположения ее масс и от раз навсегда сделанного выбора координатных осей Ox и Oy. Итак,

$$J_u = A\sin^2\varphi + 2B\sin\varphi\cos\varphi + C\cos^2\varphi.$$

Рассмотрим все возможные оси u, проходящие в плоскости пластинки через точку O. Будем откладывать на каждой из осей от точки O величину ρ, обратную корню квадратному из момента инерции J_u пластинки относительно этой оси, т. е. $\rho = \frac{1}{\sqrt{J_u}}$. Мы получим тогда

$$\frac{1}{\rho^2} = A\sin^2\varphi + 2B\sin\varphi\cos\varphi + C\cos^2\varphi.$$

Но
$$x = \rho\cos\varphi, \quad y = \rho\sin\varphi,$$
и, следовательно, уравнение этого геометрического места имеет следующий вид:

$$Cx^2 + 2Bxy + Ay^2 = 1.$$

Получается кривая 2-го порядка и притом, очевидно, конечная и замкнутая, т. е. некоторый эллипс (рис. 20), так как все остальные кривые 2-го порядка, как мы далее покажем, или бесконечные, или сводятся к одной точке.

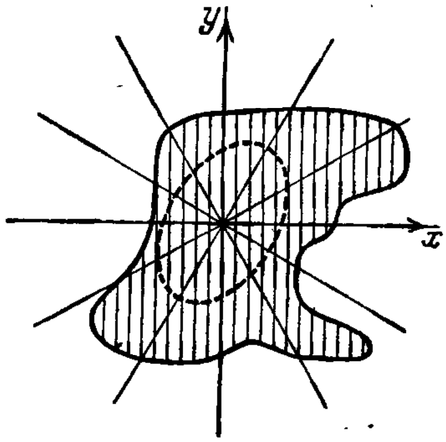


Рис. 20.

Получился замечательный результат: какова бы ни была форма и величина пластинки и расположение ее масс, величины моментов инерции ее (собственно величины ρ, обратно пропорциональные их квадратным корням) относительно различных осей, лежащих в плоскости пластинки и проходящих через заданную точку O пластинки, характеризуются некоторым эллипсом. Этот эллипс называется эллипсом инерции пластинки по отношению к точке O. Если точка O — центр тяжести пластинки, то этот эллипс называется ее центральным эллипсом инерции.

Эллипс инерции играет большую роль в механике, и особенно важное применение его имеет место в сопротивлении материалов. В сопротивлении материалов доказывается, что если мы имеем балку с каким-нибудь заданным сечением, то сопротивление ее изгибу будет пропорционально моменту инерции ее сечения относительно оси, проходящей через центр тяжести сечения и перпендикулярной к направлению изгибающей силы. Поясним это примером. Предположим, что мостик через ручей сделан из доски и доска прогибается под действием веса проходящего по ней пешехода. Если ту же доску (а не более толстую) положить «на ребро», она почти вовсе не прогнется, т. е. в положении «на ребро» доска, так сказать, прочнее. Это происходит от того, что у поперечного сечения доски, которое имеет форму довольно вытянутого прямоугольника, момент инерции этого сечения (если мыслить сечение однородно покрытым массой) относительно оси, лежащей в его плоскости, проходящей через его центр и перпендикулярной к его длинной стороне, больше, чем отно-

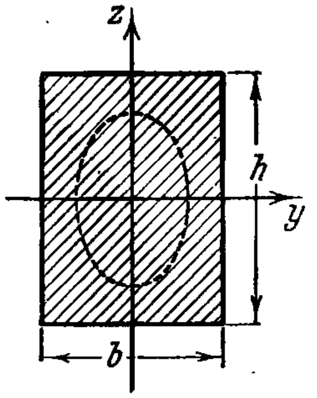


Рис. 21.

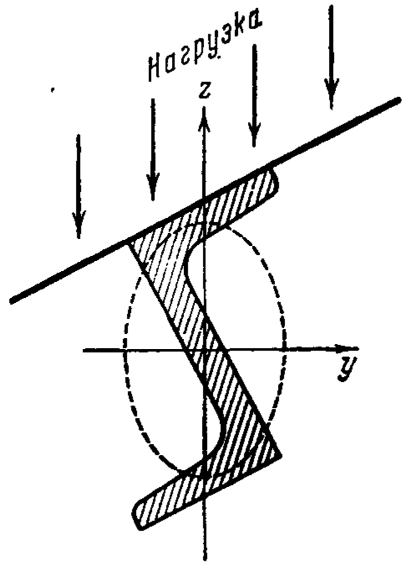


Рис. 22.

сительно оси, проходящей параллельно его длинной стороне. Если бы класть доску не в точности плашмя или на ребро, а косо, и даже если брать не доску, а брусок с любым сечением, например рельс, то все же сопротивление изгибу будет пропорционально моменту инерции этого сечения относительно соответственной оси, лежащей в его плоскости и проходящей через его центр тяжести. Жесткость при изгибе балки, таким образом, характеризуется эллипсом инерции ее сечения.

Так, для обычного прямоугольного бруса эллипс этот будет иметь вид, показанный на рис. 21. Жесткость такой балки при нагрузке в направлении оси Oz пропорциональна bh^3.

Стальные балки часто берутся ʃ-образного сечения; для такого рода балок сечение и эллипс инерции изображены на рис. 22. Наибольшей прочностью при изгибе они обладают в направлении z. При использовании их, например, для стропил крыши под нагрузкой снега и собственного веса они как раз и работают на изгиб в близком к этому наивыгоднейшему направлении.

Гипербола и ее фокальное свойство. Рассмотрим теперь уравнение

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

выражающее линию, называемую *гиперболой*. Если обозначить через c такое число, что $c^2=a^2+b^2$, то можно показать, что гипербола есть геометрическое место точек, для которых разность расстояний до двух заданных точек F_1 и F_2, лежащих на оси Ox и имеющих абсциссы c и $-c$, есть величина постоянная: $\rho_2-\rho_1=2a$ (рис. 23). Точки F_1 и F_2 и тут называют *фокусами*.

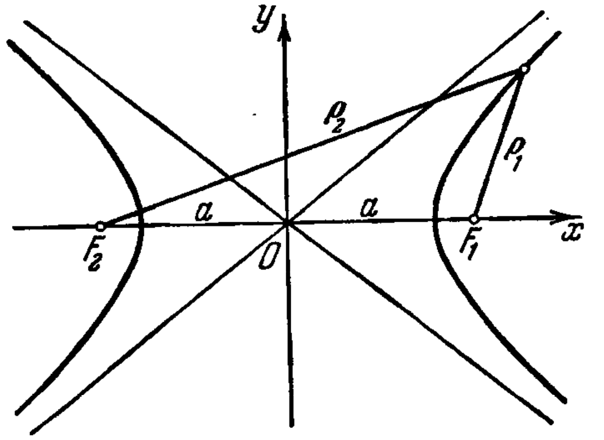


Рис. 23.

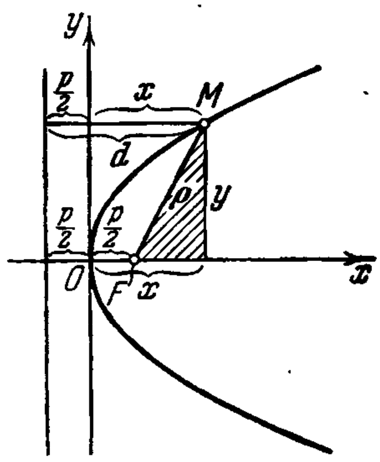


Рис. 24.

Парабола и ее директрисса. Рассмотрим, наконец, уравнение

$$y^2 = 2px$$

и будем называть линию, им выражаемую, *параболой*. Точку F, лежащую на оси Ox и имеющую абсциссу $\frac{p}{2}$, назовем *фокусом* параболы, а прямую $y = -\frac{p}{2}$, параллельную оси Oy,— *директриссой*. Пусть M — некоторая точка параболы (рис. 24) и ρ — длина ее фокального радиуса MF, а d — длина перпендикуляра, опущенного из нее на директриссу. Вычислим ρ и d для точки M. Из заштрихованного треугольника получаем $\rho^2 = \left(x - \frac{p}{2}\right)^2 + y^2$. Поскольку точка M лежит на параболе, для нее $y^2 = 2px$, следовательно,

$$\rho^2 = \left(x - \frac{p}{2}\right)^2 + 2px = \left(x + \frac{p}{2}\right)^2.$$

Но непосредственно из чертежа ясно, что $d = x + \frac{p}{2}$. Поэтому $\rho^2 = d^2$, т. е. $\rho = d$. Обратное рассуждение показывает, что если точка такова, что для нее $\rho = d$, то она лежит на рассматриваемой параболе. Парабола есть, таким образом, геометрическое место точек, расстояние ρ от каждой из которых до заданной точки F (называемой фокусом) равно расстоянию ее d до заданной прямой (называемой директриссой).

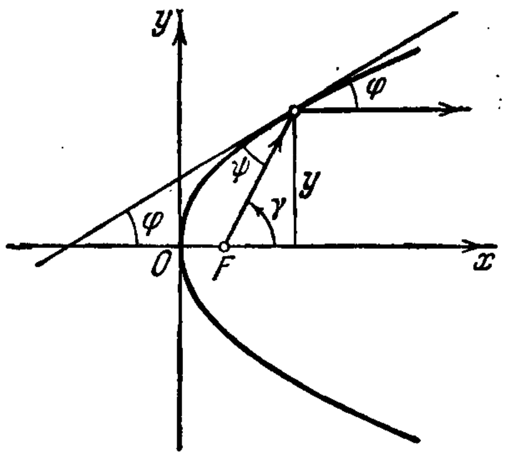


Рис. 25.

Свойство касательной к параболе. Рассмотрим одно важное свойство касательной к параболе и его приложение в оптике. Так как для параболы $y^2 = 2px$, то для нее $2y\,dy = 2p\,dx$, т. е. производная, или, что то же самое, тангенс угла φ наклона касательной к оси Ox равен $\frac{dy}{dx} = \operatorname{tg}\varphi = \frac{p}{y}$ (рис. 25).

С другой стороны, непосредственно из чертежа следует:

$$\operatorname{tg}\gamma = \frac{y}{x - \frac{p}{2}}.$$

Но

$$\operatorname{tg} 2\varphi = \frac{2\frac{p}{y}}{1 - \frac{p^2}{y^2}} = \frac{2py}{y^2 - p^2} = \frac{2py}{2px - p^2} = \frac{y}{x - \frac{p}{2}},$$

т. е. $\gamma = 2\varphi$, а так как $\gamma = \varphi + \psi$, то $\psi = \varphi$. Поэтому луч света, исходящий из фокуса F и отражающийся от элемента параболы, направление

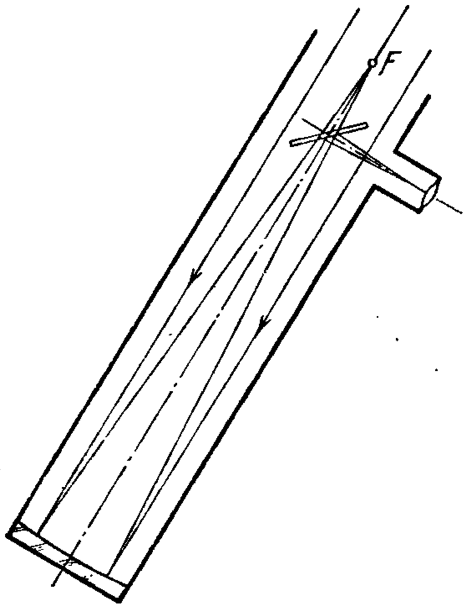


Рис. 26.

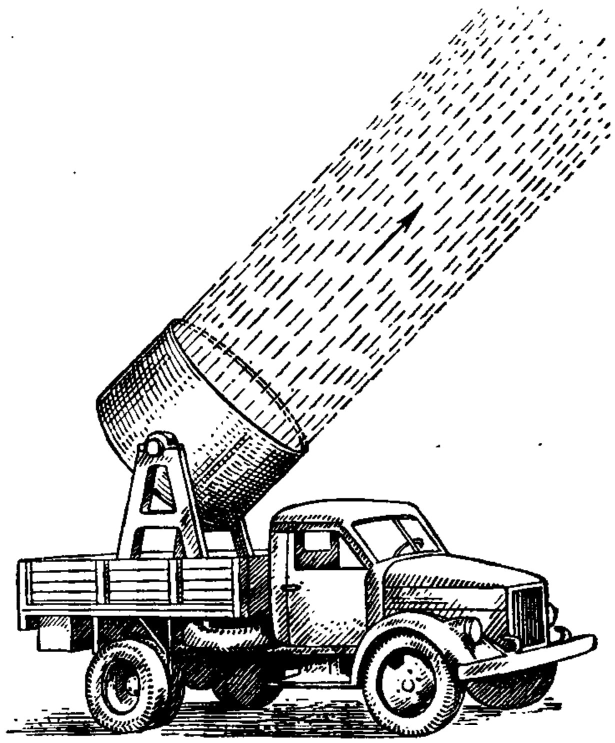


Рис. 27.

которого совпадает с направлением касательной, в силу закона: угол падения равен углу отражения, отражается дальше параллельно оси Ox, т. е. параллельно оси симметрии параболы.

На этом свойстве параболы основано изготовление отражательных телескопов, придуманных Ньютоном. Если изготовить вогнутое зеркало, поверхность которого представляет собой так называемый параболоид вращения, т. е. поверхность, получаемую от вращения параболы вокруг ее оси симметрии, то все лучи света, исходящие от какой-либо точки небесного светила, находящейся строго в направлении «оси» зеркала, соберутся зеркалом (рис. 26) в одну точку — в фокус зеркала. Лучи, идущие от какой-нибудь другой точки небесного светила, уже не совсем параллельные оси зеркала, соберутся также почти в одну точку около фокуса. Таким образом, в так называемой фокальной плоскости, т. е. плоскости, проходящей через фокус зеркала и перпендикулярной к его оси, будет получаться обратное изображение светила, и при этом чем дальше от фокуса, тем более размытое, так как лишь лучи, строго параллельные оси зеркала, собираются зеркалом в одну точку; лучи же, не совсем параллельные оси, собираются не совсем в одну точку. Получившееся изображение можно рассматривать в специальный микроскоп, так называемый окуляр телескопа, непосредственно или, чтобы не заслонять светила своей головой, повернув лучи небольшим плоским зеркалом,

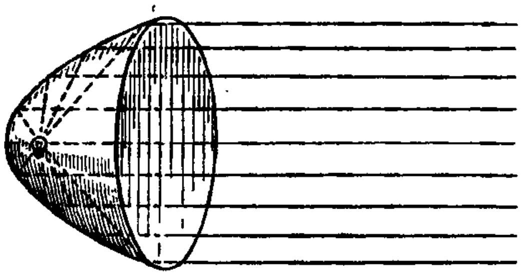


Рис. 28.

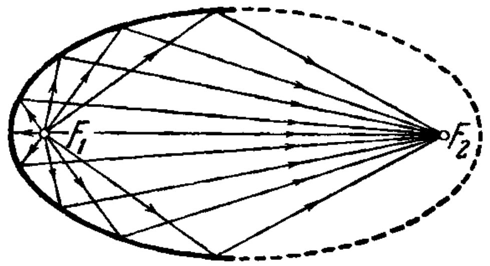


Рис. 29.

которое прикреплено в телескопе около фокуса (несколько ближе, чем фокус, к вогнутому зеркалу) под углом в 45° к оси этого зеркала.

На рассмотренном свойстве параболы основано также устройство прожекторов (рис. 27). В них, наоборот, сильный источник света помещают в фокусе параболического зеркала, и тогда лучи его, отразившись

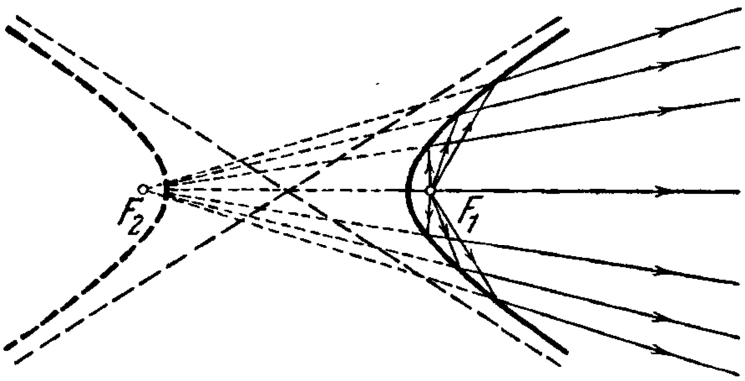


Рис. 30.

от зеркала, идут далее пучком, параллельным его оси. Так же устроены автомобильные фары (рис. 28).

В случае эллипса, как легко доказать, лучи, исходящие из одного из его фокусов F_1 и отразившиеся от эллипса, собираются в другом фокусе F_2 (рис. 29), а в случае гиперболы лучи, исходящие из одного из ее фокусов F_1 и от нее отразившиеся, как бы исходят из другого фокуса F_2 (рис. 30).

Директриссы эллипса и гиперболы. Подобно параболе, и эллипс, и гипербола имеют директриссы, притом по две директриссы. Если взять фокус и одностороннюю с ним директриссу, то для всех точек M эллипса $\frac{\rho}{d} = \varepsilon$, где ε — эксцентриситет рассматриваемого эллипса, который у эллипса всегда меньше единицы, и для всех точек соответствующей

ветви гиперболы также $\frac{\rho}{d} = \varepsilon$, где ε — эксцентриситет рассматриваемой гиперболы, причем у гиперболы он всегда больше единицы.

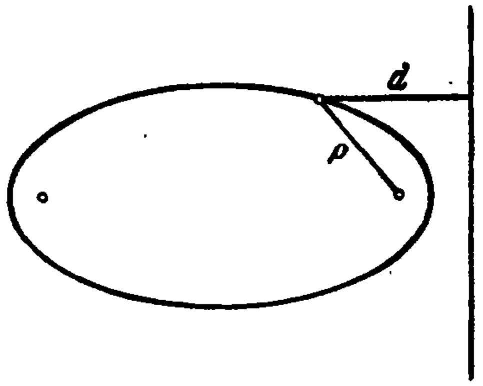


Рис. 31.

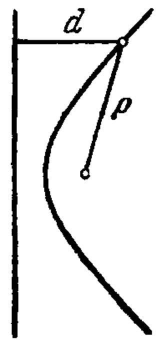


Рис. 32.

Таким образом, эллипс, парабола и одна ветвь гиперболы суть геометрические места всех тех точек плоскости, для которых отношение их расстояния ρ до фокуса к их расстоянию d до директриссы постоянно (рис. 31 и 32). Только эта постоянная у эллипсов меньше единицы, у параболы равна единице, а у гиперболы больше единицы. В этом смысле парабола есть как бы «предельный» или «переходный» случай от эллипса к гиперболе.

Конические сечения. Еще древние греки подробно исследовали линии, получаемые при пересечении прямого кругового конуса плоскостью. Если секущая плоскость образует с осью конуса угол φ в 90°, т. е. ей перпендикулярна, то в сечении получается окружность. Легко показать, что если угол φ меньше 90°, но больше угла α, который составляют с осью конуса его образующие, то получается эллипс. Если угол φ равен углу α, то получается парабола. Если угол φ меньше угла α, то в сечении получается гипербола (рис. 33).

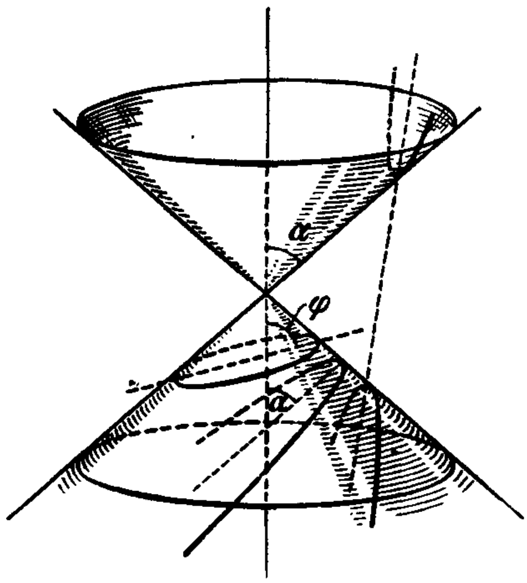


Рис. 33.

Парабола как график пропорциональности квадрату и гипербола как график обратной пропорциональности. Напомним, что график пропорциональности квадрату

$$y = kx^2$$

есть парабола (рис. 34) и что график обратной пропорциональности

$$y = \frac{k}{x}\text{, или } xy = k$$

есть гипербола (рис. 35). В последнем легко убедиться. Гипербола определялась выше как линия, выражаемая уравнением

$$\frac{x^2}{a^2}-\frac{y^2}{b^2}=1.$$

В частном же случае, когда $a=b$,— это так называемая «равносторонняя гипербола», играющая ту же роль среди гипербол, какую играет

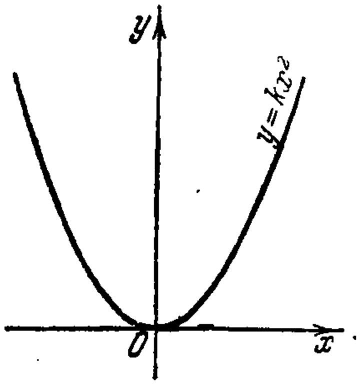
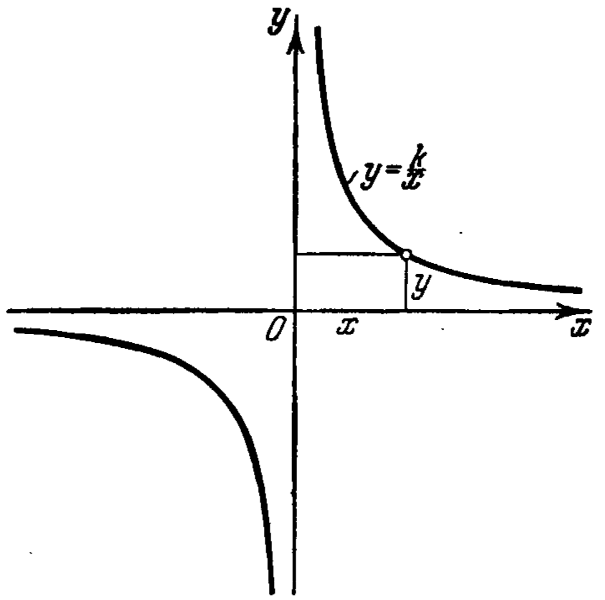


Рис. 34.



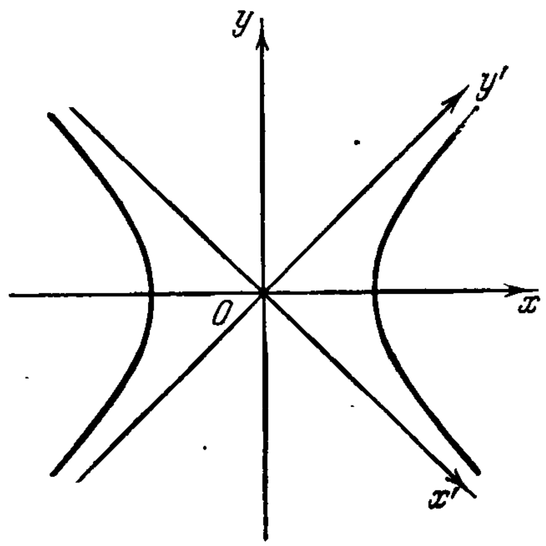


Рис. 35. Рис. 36.

окружность среди эллипсов. В этом случае, если повернуть оси координат на 45° (рис. 36), это уравнение в новых координатах (x', y') будет иметь вид

$$x'y'=k.$$

Мы сейчас рассмотрели три важнейшие линии 2-го порядка: эллипс, гиперболу и параболу, причем за определения их были взяты так называемые «канонические» уравнения

$$\frac{x^2}{a^2}+\frac{y^2}{b^2}=1,\quad \frac{x^2}{a^2}-\frac{y^2}{b^2}=1 \text{ и } y^2=2px,$$

которыми они выражаются.

Перейдем теперь к изучению общего уравнения 2-й степени с двумя переменными, а именно: к изучению вопроса о том, каковы все те линии, которые оно выражает.

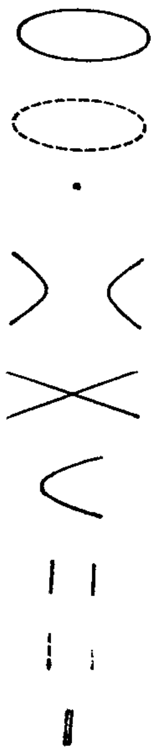
§ 8. ПРИВЕДЕНИЕ ОБЩЕГО УРАВНЕНИЯ 2-Й СТЕПЕНИ К КАНОНИЧЕСКОМУ ВИДУ

Первое последовательное изложение аналитической геометрии у Эйлера. Важным этапом в развитии аналитической геометрии было появление в 1748 г. книги «Введение в анализ», во втором томе которой впервые было дано, между прочими вещами, относившимися к теории функций и другим разделам анализа, также и изложение аналитической геометрии на плоскости с подробным исследованием линий 2-го порядка, весьма близким к даваемому в современных учебниках аналитической геометрии, а также исследованием линий высших порядков. Это был первый курс аналитической геометрии в современном смысле слова.

Идея приведения уравнения к каноническому виду. Уравнение 2-й степени[1]

$$Ax^2 + 2Bxy + Cy^2 + 2Dx + 2Ey + F = 0,$$

содержит шесть членов, а не три или только два, как рассмотренные выше канонические уравнения эллипса, гиперболы и параболы. И это не потому, что такое уравнение выражает более сложные линии, а потому, что система координат, по отношению к которой написано это уравнение, может быть к нему не приспособлена. Оказывается, что если разумно подобрать систему прямоугольных декартовых координат, то уравнение 2-й степени с двумя переменными всегда можно привести к одному из следующих канонических видов:



1)	$\frac{x^2}{a^2} + \frac{y^2}{b^2} - 1 = 0.$	Эллипс
2)	$\frac{x^2}{a^2} + \frac{y^2}{b^2} + 1 = 0.$	Мнимый эллипс
3)	$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 0.$	Точка (пара мнимых пересекающихся в действительной точке прямых)
4)	$\frac{x^2}{a^2} - \frac{y^2}{b^2} - 1 = 0.$	Гипербола
5)	$\frac{y^2}{a^2} - \frac{y^2}{b^2} = 0.$	Пара пересекающихся прямых
6)	$y^2 - 2px = 0.$	Парабола
7)	$x^2 - a^2 = 0.$	Пара параллельных прямых
8)	$x^2 + a^2 = 0.$	Пара мнимых параллельных прямых
9)	$x^2 = 0.$	Пара совпадающих прямых

где a, b, p не равны нулю.

Уравнения 1), 4) и 6) из перечисленных канонических видов нам уже известны; это канонические уравнения эллипса, гиперболы и параболы. Два из них не удовлетворяются никакими точками, а именно уравнения 2) и 8).

[1] Коэффициенты при xy, x, y мы обозначили не через B, D, E, а через $2B$, $2D$, $2E$ для упрощения записи дальнейших формул.

В самом деле, квадрат действительного числа всегда положителен или нуль, поэтому в левой части уравнения 2) уже сумма членов $\frac{x^2}{a^2}+\frac{y^2}{b^2}$ не отрицательна, кроме того, есть член $+1$, следовательно, левая часть не может равняться нулю; аналогично в уравнении 8) число x^2 не отрицательно, а a^2 положительно. Из тех же соображений следует, что уравнению 3) удовлетворяет только ($x=0$, $y=0$), т. е. одна точка — начало координат. Уравнение 5) можно написать так: $\left(\frac{x}{a}-\frac{y}{b}\right)\left(\frac{x}{a}+\frac{y}{b}\right)=0$ и тогда ясно, что ему удовлетворяют те и только те точки, для которых хотя бы одно из выражений 1-й степени $\frac{x}{a}-\frac{y}{b}$ или $\frac{x}{a}+\frac{y}{b}$ равно нулю, т. е. что линия, им выражаемая, есть совокупность этих пересекающихся прямых. Уравнение 7) аналогично дает $(x-a)(x+a)=0$, т. е. соответственная линия есть пара параллельных прямых $x=a$ и $x=-a$. Наконец, линия 9) есть частный (предельный) случай линии 7), когда $a=0$, т. е. это пара совпадающих прямых.

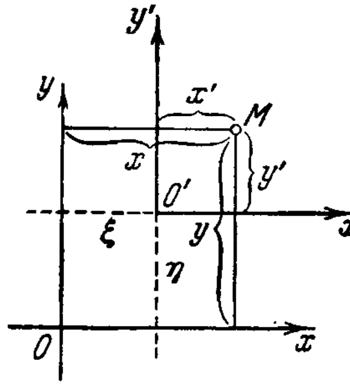


Рис. 37.

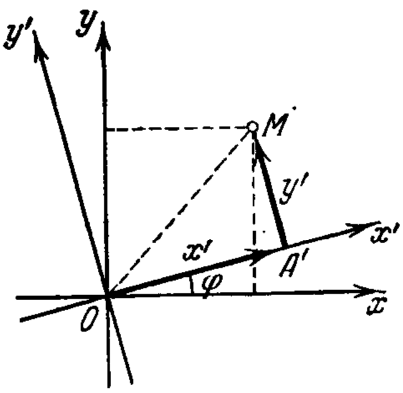


Рис. 38.

Формулы преобразования координат. Для получения указанного важного результата о возможных типах линий 2-го порядка надо сначала вывести формулы, по которым изменяются прямоугольные координаты точек при изменении координатной системы.

Пусть x, y—координаты некоторой точки M относительно осей Oxy. Перенесем эти оси параллельно самим себе в положение $O'x'y'$, и пусть координаты нового начала O' относительно старых осей суть ξ и η. Очевидно (рис. 37), что координаты x', y' точки M относительно новых осей связаны с ее координатами x, y относительно старых осей формулами

$$x=x'+\xi,$$
$$y=y'+\eta$$

— это так называемые формулы параллельного переноса осей. Если мы исходные оси Oxy повернем вокруг начала против часовой стрелки на угол φ, то, как легко вывести (рис. 38), проектируя ломаную $OA'M$, составленную новыми координатными отрезками x', y', на ось Ox и на ось Oy, получим

$$x = x' \cos \varphi - y' \sin \varphi,$$
$$y = x' \sin \varphi + y' \cos \varphi$$

— это формулы преобразования координат при повороте прямоугольной системы координат.

Если задано уравнение $F(x, y) = 0$ какой-нибудь линии относительно осей Oxy и надо записать преобразованное уравнение этой же линии, т. е. уравнение этой же линии, но относительно новых осей $O'x'y'$, то достаточно подставить в уравнение $F(x, y) = 0$ вместо x и y их выражения через x' и y', даваемые формулами преобразования. Так, например, при параллельном переносе осей получаем преобразованное уравнение $F(x' + \xi, y' + \eta) = 0$, а при повороте осей — уравнение

$$F(x' \cos \varphi - y' \sin \varphi, \quad x' \sin \varphi + y' \cos \varphi) = 0.$$

Заметим, что при преобразовании к новым осям степень уравнения не изменяется. Действительно, степень не может повыситься, так как формулы преобразования 1-й степени. Но степень не может и понизиться, так как тогда обратное преобразование координат должно было бы его повысить (а оно тоже 1-й степени).

Приведение любого уравнения 2-й степени к одному из 9 канонических видов. Покажем теперь, что каким бы ни было заданное уравнение 2-й степени с двумя переменными, всегда можно сначала так повернуть оси, а затем их параллельно перенести, что преобразованное уравнение в окончательных осях будет иметь один из видов 1), 2),...,9).

Действительно, пусть заданное уравнение 2-й степени имеет вид

$$Ax^2 + 2Bxy + Cy^2 + 2Dx + 2Ey + F = 0. \tag{8}$$

Повернем оси на некоторый угол φ, который мы сейчас подберем. Подставив в уравнение (8) вместо x и y их выражения через новые координаты (согласно формулам поворота), мы получим, собирая подобные члены, что коэффициент $2B'$ в преобразованном уравнении

$$A'x'^2 + 2B'x'y' + C'y'^2 + 2D'x' + 2E'y' + F' = 0$$

будет равен

$$2B' = -2A \sin \varphi \cos \varphi + 2B(\cos^2 \varphi - \sin^2 \varphi) + 2C \sin \varphi \cos \varphi =$$
$$= 2B \cos 2\varphi - (A - C) \sin 2\varphi.$$

Положив его равным нулю, мы получим $2B \cos 2\varphi = (A - C) \sin 2\varphi$, откуда

$$\operatorname{ctg} 2\varphi = \frac{A - C}{2B}.$$

Но так как котангенс изменяется от $-\infty$ до $+\infty$, то можно всегда найти такой угол φ, при котором удовлетворяется это равенство. При повороте осей на такой угол мы получим, что в повернутых осях $Ox'y'$ уравнение нашей линии, выражавшейся в исходных осях уравнением (8), имеет вид

$$A'x'^2 + C'y'^2 + 2D'x' + 2E'y' + F = 0, \tag{9}$$

т. е. что оно уже не содержит члена с произведением координат. (F остается прежним, так как формулы поворота не содержат постоянных членов.)

Теперь перенесем уже повернутые оси $Ox'y'$ параллельно самим себе в положение $O''x''y''$, и пусть координаты нового начала O'' относительно осей $Ox'y'$ суть ξ', η'. Уравнение нашей линии в этих окончательных осях будет

$$A'(x'' + \xi')^2 + C'(y'' + \eta')^2 + 2D'(x'' + \xi') + 2E'(y'' + \eta') + F = 0. \tag{10}$$

Покажем, что можно всегда так подобрать ξ' и η', т. е. так перенести оси $Ox'y'$ параллельно самим себе, чтобы окончательное уравнение в осях $O''x''y''$ имело один из канонических видов 1), 2), ..., 9).

Раскрывая все скобки в уравнении (10) и делая приведение подобных членов, получаем

$$A'x''^2 + C'y''^2 + 2(A'\xi' + D')x'' + 2(C'\eta' + E')y'' + F' = 0, \tag{10'}$$

где мы через F' обозначили сумму всех постоянных членов; какова она, нас это сейчас не интересует.

Рассмотрим три возможных случая.

I. A' и C' оба не равны нулю. В этом случае, взяв $\xi' = -\frac{D'}{A'}$, $\eta' = -\frac{E'}{C'}$, мы уничтожим члены с первыми степенями x'' и y'' и получим уравнение вида

$$A'x''^2 + C'y''^2 + F' = 0. \tag{I}$$

II. $A' \neq 0$, $C' = 0$, но $E' \neq 0$. Взяв $\xi' = -\frac{D'}{A'}$, $\eta' = 0$, т. е. $y'' = y'$, мы получим уравнение

$$A'x''^2 + 2E'y' + F' = 0,$$

или

$$A'x''^2 + 2E'\left(y' + \frac{F'}{2E'}\right) = 0.$$

Сделав еще параллельный перенос вдоль оси Oy' на величину $\eta'' = -\frac{F'}{2E'}$, мы найдем, что $y' = y'' - \frac{F'}{2E'}$, т. е. $y' + \frac{F'}{2E'} = y''$, т. е. получим уравнение

$$A'x''^2 + 2E'y'' = 0. \tag{II}$$

Если бы было $A' = 0$, $C' \neq 0$ и $D' \neq 0$, мы просто поменяли бы ролями x и y и получили бы этот же случай.

III. $A' \neq 0$, $C' = 0$, $E' = 0$. Взяв опять $\xi' = -\frac{D'}{A'}$, $\eta' = 0$, мы получим уравнение

$$A'x''^2 + F' = 0. \tag{III}$$

Если было бы $A' = 0$, $C' \neq 0$, $D' = 0$, мы опять поменяли бы ролями x и y.

Тут перебраны все возможности, так как одновременно и A' и C' равными нулю быть не могут, поскольку тогда понизилась бы степень уравнения, а мы видели, что при преобразованиях координат она не изменяется.

Итак, соответственным выбором прямоугольных координат всякое уравнение 2-й степени можно привести к одному из так называемых трех «приведенных» уравнений: (I), (II) или (III).

Пусть уравнение имеет вид (I) (в этом случае A' и C' не равны нулю). Если $F' \neq 0$, то, написав уравнение (I) так:

$$\frac{x''^2}{-\frac{F'}{A'}} + \frac{y''^2}{-\frac{F'}{C'}} - 1 = 0,$$

мы, в зависимости от знаков A', C', F', приходим к одному из уравнений 1), 2) или 4). Если знаменатель при x''^2 отрицателен, а при y''^2 положителен, то ~~надо еще изменить~~ название оси $O''x''$ на $O''y''$ и оси $O''y''$ на $O''x''$.

Если $F' = 0$, то, написав уравнение (I) в виде

$$\frac{x''^2}{\frac{1}{A'}} + \frac{y''^2}{\frac{1}{C'}} = 0,$$

мы приходим к уравнениям 3) или 5).

Если уравнение будет вида (II) (в каковом случае A' и E' оба не равны нулю), то, написав его так:

$$x''^2 + \frac{2E'}{A'}y'' = 0,$$

и обозначив $-\frac{E'}{A'}$ через p и изменив названия осей $O''x''$ и $O''y''$ на $O''y''$ и $O''x''$, получим уравнение 6).

Если, наконец, мы имеем уравнение вида (III) (причем тогда $A' \neq 0$), то его можно переписать так: $x''^2 + \frac{F'}{A'} = 0$, и получится одно из уравнений 7), 8) или 9).

Только что доказанная важная теорема о возможности приведения всякого уравнения 2-й степени к одному из 9 канонических видов была уже подробно разобрана Эйлером. Рассуждения в книге Эйлера лишь по форме отличаются от сейчас приведенных.

§ 9. ЗАДАНИЕ СИЛ, СКОРОСТЕЙ И УСКОРЕНИЙ ТРОЙКАМИ ЧИСЕЛ. ТЕОРИЯ ВЕКТОРОВ

Следующий после Эйлера важный шаг был сделан Лагранжем. В своей «Аналитической механике» 1788 г. Лагранж арифметизировал силы скорости и ускорения совершенно аналогично тому, как Декарт арифметизировал точки. Эта идея Лагранжа, проводимая в его книге, впоследствии, в форме так называемой теории векторов, оказалась важнейшим подспорьем в физике, механике и технике.

Прямоугольные координаты в пространстве. Заметим прежде всего, что ни Декарт, ни Ньютон не разрабатывали аналитической геометрии в пространстве. Это было сделано уже позже, в первой половине XVIII в., Клеро и Лагиром. Для задания точки M в пространстве выбирают три взаимно перпендикулярные оси Ox, Oy и Oz и рассматривают (рис. 39) численные величины расстояний точки M от плоскостей Oyz, Oxz и Oxy, взятые с соответствующими знаками, так называемые *абсциссу* x, *ординату* y и *аппликату* z точки M.

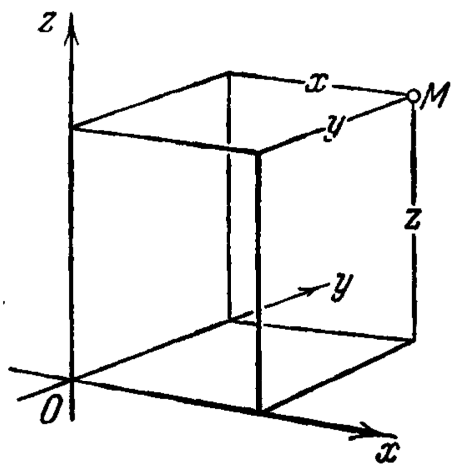


Рис. 39.

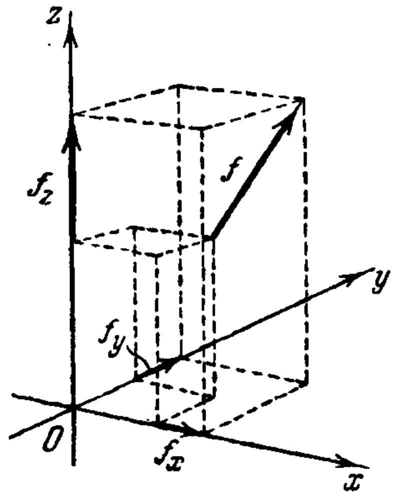


Рис. 40.

Арифметизация сил, скоростей и ускорений, введенная Лагранжем. Рассмотрим (рис. 40) некоторую силу f, которую можно в условных единицах изобразить отрезком со стрелкой, имеющим определенную длину и направление. Лагранж говорит, что силу эту f можно разложить на три составляющие: f_x, f_y и f_z, идущие соответственно по оси Ox, Oy и Oz; эти составляющие, как направленные отрезки на осях, можно уже задать просто числами, положительными или отрицательными в зависимости от того, направлена ли составляющая в положительном направлении оси или ему противоположном. Таким образом можно, например, рассматривать силу (2, 3, 4) или силу (1, —2, 5) и т. д. При сложении сил по правилу параллелограмма соответствующие их составляющие, как легко показать (ниже это будет сделано), тоже складываются. Например, сумма написанных сил есть сила $(2+1,\ 3-2,\ 4+5)=(3,\ 1,\ 9)$. То же самое можно делать для скоростей и ускорений. Во всех задачах механики можно все уравнения механики, связывающие силы, скорости и ускорения, писать уже как уравнения, связывающие их составляющие, т. е. связывающие уже просто числа; только каждое уравнение придется писать в виде трех уравнений: одно для x-ов, другое для y-ов, третье для z-ов.

Только через сто лет после Лагранжа математики и физики, особенно под влиянием развивавшейся тогда теории электричества, начали широко рассматривать общую теорию таких отрезков, имеющих определенную длину и направление. Такие отрезки были названы векторами.

Теория векторов имеет большое значение в механике, физике и технике, а ее алгебраическая часть, так называемая алгебра векторов (в отличие от векторного анализа), является сейчас существенной составной частью аналитической геометрии.

Алгебра векторов. Любой направленный отрезок (все равно, что он представляет: силу, скорость, ускорение или еще что-нибудь), т. е.

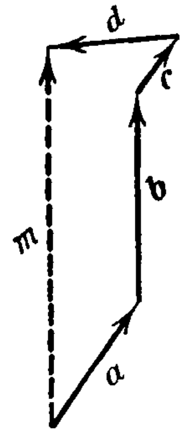


Рис. 41.

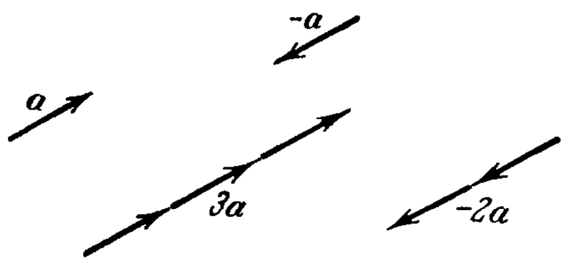


Рис. 42.

отрезок, имеющий данную длину и определенное направление, называют *вектором*. Два вектора называют *равными*, если они имеют одну и ту же длину и одно и то же направление, т. е. в самом понятии «вектор» учитывается только его длина и его направление. Векторы можно складывать. Пусть даны векторы $\boldsymbol{a}, \boldsymbol{b}, \ldots, \boldsymbol{d}$. Отложим от некоторой точки вектор $\boldsymbol{a}$, затем от его конца вектор $\boldsymbol{b}$ и т. д. Мы получим так называемую векторную ломаную $\boldsymbol{ab} \ldots \boldsymbol{d}$ (рис. 41). Вектор $\boldsymbol{m}$, начало которого совпадает с началом первого вектора $\boldsymbol{a}$ этой ломаной, а конец — с концом последнего вектора $\boldsymbol{d}$ этой ломаной, называют *суммой* этих векторов

$$\boldsymbol{m} = \boldsymbol{a} + \boldsymbol{b} + \ldots + \boldsymbol{d}. \tag{11}$$

Легко доказать, что вектор $\boldsymbol{m}$ не зависит от того, в каком порядке брать слагаемые $\boldsymbol{a}, \boldsymbol{b}, \ldots, \boldsymbol{d}$.

Вектор, равный по длине вектору $\boldsymbol{a}$, но обратного ему направления, называется *обратным* ему вектором и обозначается через $-\boldsymbol{a}$.

Вычитанием вектора $\boldsymbol{a}$ называют прибавление обратного ему вектора.

Обыкновенные действительные числа в векторном исчислении принято называть скалярными. Пусть задан вектор $\boldsymbol{a}$ (рис. 42) и скаляр λ, тогда произведением вектора $\boldsymbol{a}$ на скаляр (число) λ, т. е. $\lambda\boldsymbol{a}$, называют вектор, длина которого равна произведению длины $|\boldsymbol{a}|$ вектора $\boldsymbol{a}$ на абсолютную величину $|\lambda|$ числа λ, а направление такое же, как у $\boldsymbol{a}$, если $\lambda > 0$, и обратное, если $\lambda < 0$.

Рассмотрим систему прямоугольных декартовых координат $Oxyz$ и векторы $\boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{e}_3$, имеющие длины, равные единице, и направления, совпадающие с положительными направлениями осей Ox, Oy, Oz. Очевидно, что до любой взятой точки M (рис. 43) пространства можно дойти из начала O столько-то «раз» (целое, дробное или иррациональное, положительное или отрицательное «число раз») пройдя вектор $\boldsymbol{e}_1$, затем столько-то «раз» пройдя вектор $\boldsymbol{e}_2$, наконец, столько-то «раз», пройдя

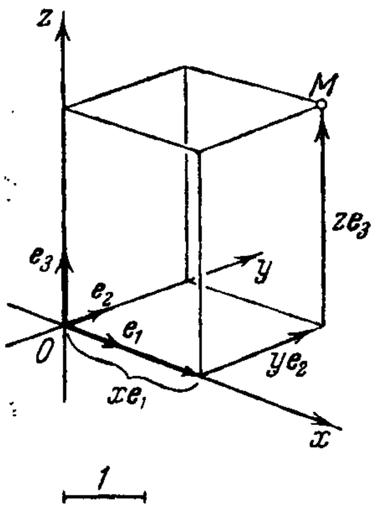


Рис. 43.

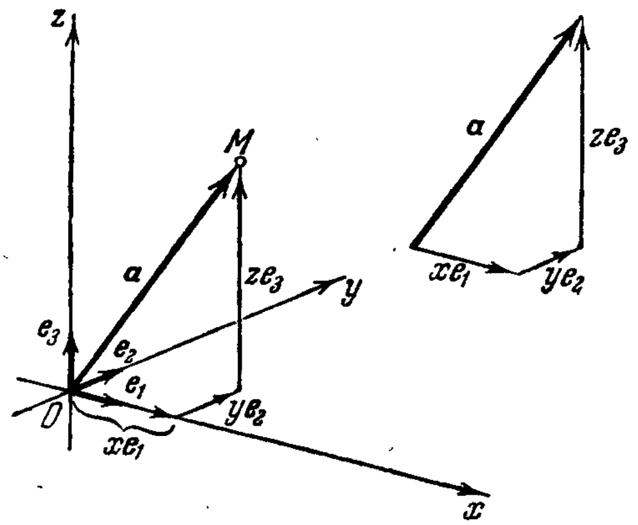


Рис. 44.

вектор $\boldsymbol{e}_3$. Числа x, y, z, показывающие, сколько «раз» при этом надо пройти векторы $\boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{e}_3$, суть, очевидно, просто декартовы координаты точки M.

Пусть дан некоторый вектор $\boldsymbol{a}$; будем двигать точку от его начала к его концу и движение это разложим на движения параллельно осям Ox, Oy и Oz, тогда, если придется при этом точку передвинуть на $x\boldsymbol{e}_1$ параллельно оси Ox, на $y\boldsymbol{e}_2$ параллельно оси Oy и на $z\boldsymbol{e}_3$ параллельно оси Oz, то

$$\boldsymbol{a} = x\boldsymbol{e}_1 + y\boldsymbol{e}_2 + z\boldsymbol{e}_3. \qquad (12)$$

Числа x, y, z называются *координатами* вектора $\boldsymbol{a}$. Это, очевидно, просто координаты конца M этого вектора, если его начало поместить в начале координат O (рис. 44). Отсюда ясно, что при сложении векторов складываются их одноименные координаты, а при вычитании — вычитаются. Если первый вектор «уносит» вдоль оси Ox на $x\boldsymbol{e}_1$, а второй — на $x'\boldsymbol{e}_1$, то их сумма, очевидно, «уносит» вдоль оси Ox на $(x+x')\,\boldsymbol{e}_1$ и т. д. (рис. 45). Ясно также, что при умножении вектора на число его координаты умножаются на это число.

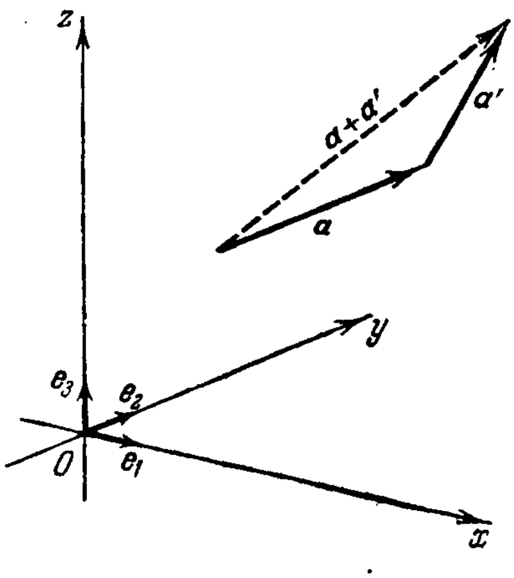


Рис. 45.

Скалярное произведение и его свойства. Если даны два вектора $\boldsymbol{a}$ и $\boldsymbol{b}$, то число, равное произведению их длин на косинус угла между ними $|\boldsymbol{a}||\boldsymbol{b}|\cos\varphi$, называют их *скалярным произведением* и обозначают через $\boldsymbol{ab}$ или $(\boldsymbol{ab})$. Пусть (x, y, z)—координаты вектора $\boldsymbol{a}$, а $\bar{x}$, $\bar{y}$, $\bar{z}$ — координаты вектора $\boldsymbol{b}$, в таком случае скалярное произведение равно

$$\boldsymbol{ab} = x\bar{x} + y\bar{y} + z\bar{z}, \tag{13}$$

т. е. равно сумме произведений их одноименных координат.

Эта важная формула доказывается так. Сделаем сперва следующие замечания:

1° Если мы помножим один из скалярно умножаемых векторов, например $\boldsymbol{a}$, на число λ, то на это же число помножится, очевидно, и их скалярное произведение, т. е.

$$(\lambda\boldsymbol{a})\,\boldsymbol{b} = \lambda\,(\boldsymbol{ab}).$$

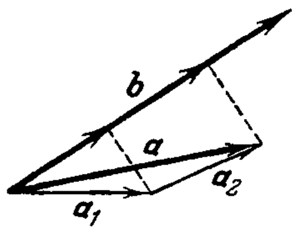


Рис. 46.

2° Скалярное умножение распределительно, т. е. если вектор $\boldsymbol{a} = \boldsymbol{a}_1 + \boldsymbol{a}_2$, то $\boldsymbol{ab} = \boldsymbol{a}_1\boldsymbol{b} + \boldsymbol{a}_2\boldsymbol{b}$.

Действительно, левая часть этого равенства равна произведению длины вектора $\boldsymbol{b}$ на численную величину проекции вектора $\boldsymbol{a}$ на ось вектора $\boldsymbol{b}$ (рис. 46), а правая — произведению длины вектора $\boldsymbol{b}$ на сумму численных величин проекции вектора $\boldsymbol{a}_1$ и $\boldsymbol{a}_2$ на ось вектора $\boldsymbol{b}$. Но пр. $\boldsymbol{a}$ = пр. $\boldsymbol{a}_1$ + пр. $\boldsymbol{a}_2$, что и доказывает справедливость равенства.

Пусть теперь даны два вектора $\boldsymbol{a}$ и $\boldsymbol{b}$, и их разложения по векторам $\boldsymbol{e}_1$, $\boldsymbol{e}_2$, $\boldsymbol{e}_3$ суть $\boldsymbol{a} = x\boldsymbol{e}_1 + y\boldsymbol{e}_2 + z\boldsymbol{e}_3$, $\boldsymbol{b} = \bar{x}\boldsymbol{e}_1 + \bar{y}\boldsymbol{e}_2 + \bar{z}\boldsymbol{e}_3$, тогда

$$\boldsymbol{ab} = (x\boldsymbol{e}_1 + y\boldsymbol{e}_2 + z\boldsymbol{e}_3)(\bar{x}\boldsymbol{e}_1 + \bar{y}\boldsymbol{e}_2 + \bar{z}\boldsymbol{e}_3).$$

В силу распределительности (2°) скалярного умножения суммы векторов, стоящих в скобках, можно умножать как многочлены, а в силу (1°) скалярные множители можно в каждом полученном члене собирать в начале, поэтому

$$\boldsymbol{ab} = x\bar{x}\boldsymbol{e}_1\boldsymbol{e}_1 + x\bar{y}\boldsymbol{e}_1\boldsymbol{e}_2 + x\bar{z}\boldsymbol{e}_1\boldsymbol{e}_3 + y\bar{x}\boldsymbol{e}_2\boldsymbol{e}_1 + y\bar{y}\boldsymbol{e}_2\boldsymbol{e}_2 + y\bar{z}\boldsymbol{e}_2\boldsymbol{e}_3 + \\ + z\bar{x}\boldsymbol{e}_3\boldsymbol{e}_1 + z\bar{y}\boldsymbol{e}_3\boldsymbol{e}_2 + z\bar{z}\boldsymbol{e}_3\boldsymbol{e}_3.$$

Но

$$|\boldsymbol{e}_1| = |\boldsymbol{e}_2| = |\boldsymbol{e}_3| = 1, \quad \cos 0 = 1 \text{ и } \cos 90° = 0.$$

Следовательно

$$\begin{aligned} \boldsymbol{e}_1\boldsymbol{e}_1 &= 1, & \boldsymbol{e}_1\boldsymbol{e}_2 &= 0, & \boldsymbol{e}_1\boldsymbol{e}_3 &= 0, \\ \boldsymbol{e}_2\boldsymbol{e}_1 &= 0, & \boldsymbol{e}_2\boldsymbol{e}_2 &= 1, & \boldsymbol{e}_2\boldsymbol{e}_3 &= 0, \\ \boldsymbol{e}_3\boldsymbol{e}_1 &= 0, & \boldsymbol{e}_3\boldsymbol{e}_2 &= 0, & \boldsymbol{e}_3\boldsymbol{e}_3 &= 1. \end{aligned}$$

Таким образом,

$$\boldsymbol{ab} = x\bar{x} + y\bar{y} + z\bar{z}. \tag{14}$$

Заметим, в частности, что если векторы $\boldsymbol{a}$ и $\boldsymbol{b}$ взаимно перпендикулярны, то $\varphi = 90°$ и $\cos\varphi = 0$. Поэтому равенство

$$x\bar{x} + y\bar{y} + z\bar{z} = 0 \tag{15}$$

служит легко проверяемым условием перпендикулярности векторов $\boldsymbol{a}$ и $\boldsymbol{b}$.

Угол между двумя направлениями. Рассмотрим некоторое направление, характеризуемое своими углами α, β, γ с осями координат. Проведем

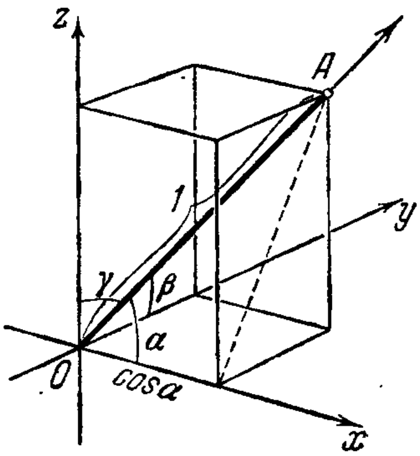


Рис. 47.

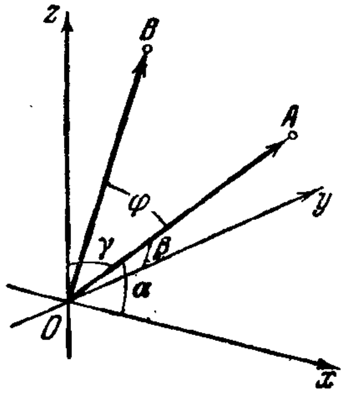


Рис. 48.

через начало координат прямую этого направления и отложим на ней от начала отрезок OA, равный единице (рис. 47). В таком случае координаты точки A, т. е. координаты вектора $\overrightarrow{OA}$, суть как раз $\cos\alpha$, $\cos\beta$ и $\cos\gamma$. Если имеется еще второе направление, задаваемое углами $\bar{\alpha}$, $\bar{\beta}$, $\bar{\gamma}$, то аналогичный вектор $\overrightarrow{OB}$ для этого второго направления имеет координаты $\cos\bar{\alpha}$, $\cos\bar{\beta}$, $\cos\bar{\gamma}$ (рис. 48). Пусть φ — угол между этими векторами, тогда скалярное их произведение равно $1 \cdot 1 \cos\varphi$, откуда мы находим

$$\cos\varphi = \cos\alpha\cos\bar{\alpha} + \cos\beta\cos\bar{\beta} + \cos\gamma\cos\bar{\gamma}. \tag{16}$$

Это весьма важная формула для косинуса угла между двумя направлениями.

§ 10. АНАЛИТИЧЕСКАЯ ГЕОМЕТРИЯ В ПРОСТРАНСТВЕ. УРАВНЕНИЕ ПОВЕРХНОСТИ В ПРОСТРАНСТВЕ И УРАВНЕНИЯ ЛИНИИ

Если дано уравнение $z = f(x, y)$ и если считать x и y абсциссой и ординатой, а z — апликатой точки, то это уравнение выражает собой некоторую поверхность P, которую можно получить, восставляя над точками (x, y) плоскости Oxy перпендикуляры, равные z. Геометрическое место концов этих перпендикуляров дает поверхность P, выражаемую этим уравнением. Если уравнение, связывающее x, y и z, не разрешено относительно z, то его можно разрешить относительно z и затем построить эту поверхность P. Вообще в аналитической геометрии поверхностью, выражаемой уравнением с тремя переменными x, y, z, называется совокупность всех точек пространства, координаты которых x, y, z удовлетворяют этому уравнению (рис. 49).

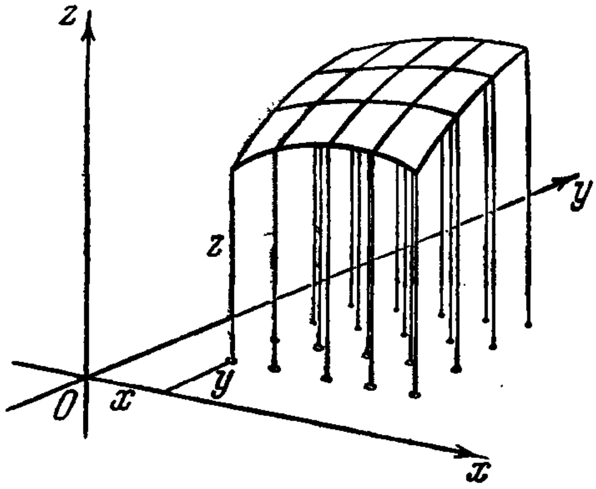


Рис. 49.

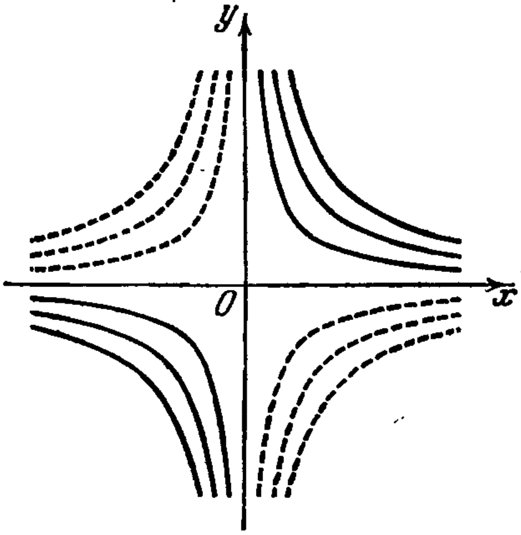


Рис. 50.

Функцию двух переменных $f(x, y)$, как уже говорилось в главе II можно изобразить не только поверхностью P, но и системой ее линий уровня, т. е. линий в плоскости Oxy, на каждой из которых функция

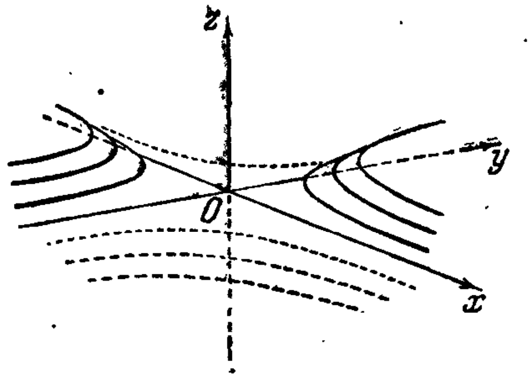


Рис. 51.

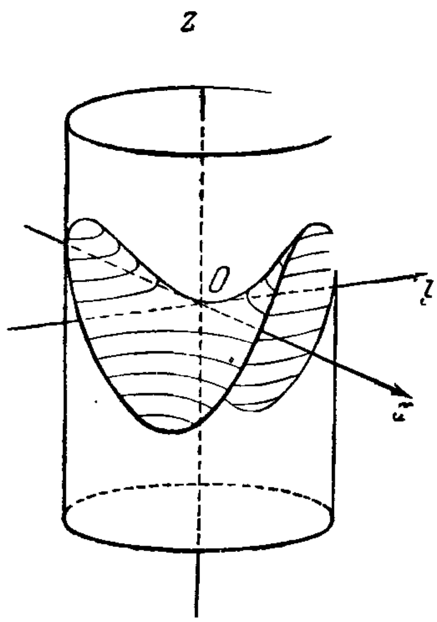


Рис. 52.

$f(x, y)$ имеет постоянное значение. Эта система линий будет, очевидно не чем иным, как топографической картой поверхности P на плоскостя Oxy.

Пример. Уравнение $xy = z$ дает, например, такие линии уровня $\ldots, xy = -3,\ xy = -2,\ xy = -1,\ xy = 0,\ xy = 1;\ xy = 2,\ xy = 3, \ldots$ Все они (рис. 50) — гиперболы, кроме линии $xy = 0$, которая представляет собою крест осей координат. Получается, очевидно, некоторая седлообразная поверхность (рис. 51) (так называемый гиперболический параболоид).

Чтобы задать линию в пространстве, можно задать уравнения каких-либо двух поверхностей P и Q, которые по ней пересекаются. Например, система

$$xy = z,$$
$$x^2 + y^2 = 1$$

задает пространственную линию (рис. 52). Уравнение $xy = z$ определяет рассмотренный сейчас гиперболический параболоид, а уравнение $x^2 + y^2 = 1$ — круглый цилиндр с радиусом, равным единице, осью которого является ось Oz. Система рассматриваемых уравнений задает, следовательно, линию пересечения параболоида с цилиндром, что и изображено на рис. 52.

Если в этой системе одну из неизвестных, например x, выбирать по произволу, а затем решать эту систему по отношению к y и z, то будут получаться координаты x, y, z разных точек этой линии.

Уравнение плоскости и уравнения прямой. Можно показать, что всякое уравнение 1-й степени с тремя переменными

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

выражает некоторую плоскость, и обратно. В силу сказанного очевидно, что прямую можно задавать системой двух таких уравнений:

$$A_1x + B_1y + C_1z + D_1 = 0,$$
$$A_2x + B_2y + C_2z + D_2 = 0,$$

т. е. как линию пересечения двух плоскостей.

Общее уравнение 2-й степени с тремя переменными и 17 его канонических видов. Уравнение 2-й степени с тремя переменными

$$A_1x^2 + A_2y^2 + A_3z^2 + 2B_1yz + 2B_2xz + 2B_3xy + 2C_1x + 2C_2y + 2C_3z + D = 0, \quad (17)$$

содержит 10 членов. Аналогично тому, как это было сделано для уравнения с двумя переменными, можно показать, что, соответственно повернув заданные прямоугольные координаты вокруг начала, можно привести уравнение (17) к виду

$$A_1'x'^2 + A_2'y'^2 + A_3'z'^2 + 2C_1'x' + 2C_2'y' + 2C_3'z' + D = 0, \quad (18)$$

т. е. уничтожить члены с произведениями переменных. Однако здесь доказательство возможности так упростить уравнение гораздо труднее, чем в случае плоскости. Трудность доказательства объясняется тем, что на плоскости поворот вокруг точки задается одним углом φ, который мы и подбирали. В пространстве же поворот тела вокруг неподвижной точки задается тремя независимыми углами (углы Эйлера) φ, θ, ψ, и притом довольно сложно. Поэтому приходится доказывать возможность избавления уравнения от членов с произведениями переменных обходным путем.

(см. в главе XVI (том 3) теорию приведения при помощи ортогонального преобразования квадратичной формы к сумме квадратов). Далее, так же как на плоскости, производится еще тот или иной параллельный перенос осей, и уравнение упрощается, после чего уравнение (18) окончательно принимает один из следующих канонических видов:

1) $\frac{x^2}{a^2}+\frac{y^2}{b^2}+\frac{z^2}{c^2}-1=0$		Эллипсоид
2) $\frac{x^2}{a^2}+\frac{y^2}{b^2}+\frac{z^2}{c^2}+1=0$		Мнимый эллипсоид
3) $\frac{x^2}{a^2}+\frac{y^2}{b^2}-\frac{z^2}{c^2}-1=0$		Однополостный гиперболоид
4) $\frac{x^2}{a^2}+\frac{y^2}{b^2}-\frac{z^2}{c^2}+1=0$		Двуполостный гиперболоид
5) $\frac{x^2}{a^2}+\frac{y^2}{b^2}-\frac{z^2}{c^2}=0$		Конус 2-го порядка
6) $\frac{x^2}{a^2}+\frac{y^2}{b^2}+\frac{z^2}{c^2}=0$		Мнимый конус 2-го порядка
7) $\frac{x^2}{a^2}+\frac{y^2}{b^2}-2cz=0$		Эллиптический параболоид
8) $\frac{x^2}{a^2}-\frac{y^2}{b^2}-2cz=0$		Гиперболический параболоид
9) $\frac{x^2}{a^2}+\frac{y^2}{b^2}-1=0$		Эллиптический цилиндр
10) $\frac{x^2}{a^2}+\frac{y^2}{b^2}+1=0$		Мнимый эллиптический цилиндр
11) $\frac{x^2}{a^2}+\frac{y^2}{b^2}=0$		Пара мнимых пересекающихся плоскостей
12) $\frac{x^2}{a^2}-\frac{y^2}{b^2}-1=0$		Гиперболический цилиндр
13) $\frac{x^2}{a^2}-\frac{y^2}{b^2}=0$		Пара пересекающихся плоскостей
14) $y^2-2px=0$		Параболический цилиндр
15) $x^2-a^2=0$		Пара параллельных плоскостей
16) $x^2+a^2=0$		Пара мнимых параллельных плоскостей
17) $x^2=0$		Пара совпадающих плоскостей

Последние девять канонических уравнений (9—17) не содержат членов с z и представляют собой как раз канонические уравнения линий 2-го порядка на плоскости Oxy. В пространстве эти уравнения выражают цилиндры, направляющие которых суть соответственные линии 2-го порядка в плоскости Oxy и образующие которых параллельны оси Oz. Действительно, если какому-нибудь из этих уравнений удовлетворяет точка с координатами $(x_1, y_1, 0)$, то ему же удовлетворяет и точка с координатами (x_1, y_1, z), какое бы ни взять z, так как все равно в уравнении нет членов с z.

Из уравнений 1)—8), как легко видеть, уравнению 2) не удовлетворяет никакая точка с действительными x, y, z, а уравнению 6) удовлетворяет только одна такая точка (0, 0, 0) — начало координат. Остается, таким образом, изучить только шесть уравнений: 1), 3), 4), 5), 7), 8).

Эллипсоид. Сравним поверхности, выражаемые уравнением $\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} - 1 = 0$ и уравнением $x^2 + y^2 + z^2 - 1 = 0$. Второе из этих уравнений, очевидно, представляет собой уравнение сферы C с центром в начале координат и радиусом, равным единице, так как $x^2 + y^2 + z^2$ есть квадрат расстояния от точки (x, y, z) до начала O. Если (x, y, z) — точка, лежащая на сфере, т. е. удовлетворяющая второму уравнению, то (ax, by, cz) — точка, удовлетворяющая своими координатами первому уравнению. Поверхность, выражаемая первым уравнением, получается, таким образом, из сферы C, если все абсциссы x точек сферы заменить на ax, y — на by, z — на cz, т. е. если сферу C равномерно растянуть от плоскостей Oyz, Oxz и Oxy с коэффициентами растяжения a, b и c. Поверхность эта называется эллипсоидом (рис. 53).

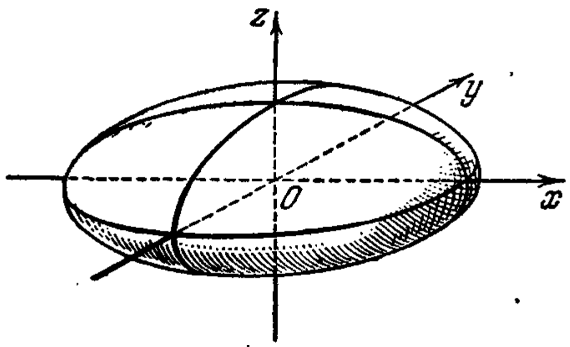


Рис. 53.

Гиперболоиды и конус 2-го порядка. Рассмотрим теперь уравнения 3), 4) и 5), т. е. уравнение вида

$$\frac{x^2}{a^4} + \frac{y^2}{b^2} - \frac{z^2}{c^2} = \delta, \tag{19}$$

где $\delta = 1$, -1 или 0. Сравним зто уравнение с уравнением

$$\frac{x^2}{a^2} + \frac{y^2}{a^2} - \frac{z^2}{c^2} = \delta, \tag{20}$$

у которого знаменатель при y^2 тоже a^2, а не b^2, как в уравнении (19). Аналогично предыдущему замечаем, что поверхность (19) получается из поверхности (20) в результате растяжения ее от плоскости Oxz с коэффициентом b/a.

Посмотрим теперь, что представляет собою поверхность (20). Возьмем какую-нибудь плоскость $z = h$, перпендикулярную к оси Oz, и исследуем, как она пересекает поверхность (20). Подставляя $z = h$ в уравнение (20), мы получим уравнение

$$x^2 + y^2 = a^2\left(\delta + \frac{h^2}{c^2}\right).$$

Если $\delta + \frac{h^2}{c^2}$ положительно, то полученное уравнение, вместе с $z = h$, дает окружность, лежащую в плоскости $z = h$ с центром на оси Oz. Если $\delta + \frac{h^2}{c^2}$ отрицательно, что может быть только при $\delta = -1$ и малых h^2, то плоскость $z = h$ вовсе не пересекает поверхности (20), так как сумма квадратов $x^2 + y^2$ не может быть отрицательным числом.

Вся поверхность (20) состоит, таким образом, из окружностей, лежащих в плоскостях, перпендикулярных к оси Oz и имеющих свои центры

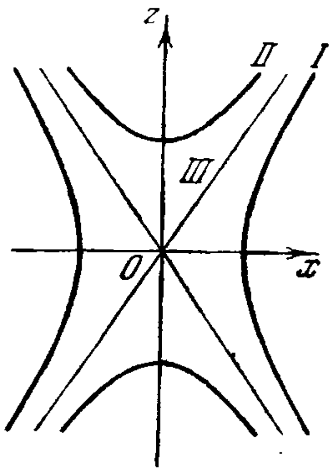


Рис. 54.

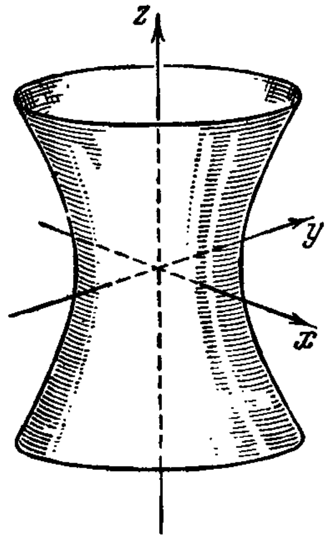


Рис. 55.

на оси Oz. Но в таком случае поверхность (20) — это поверхность вращения вокруг оси Oz. Надо только пересечь ее какой-нибудь плоскостью, про-

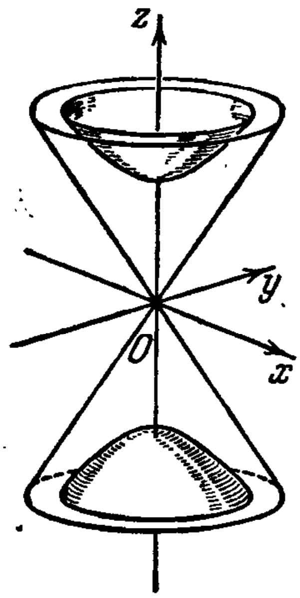


Рис. 56.

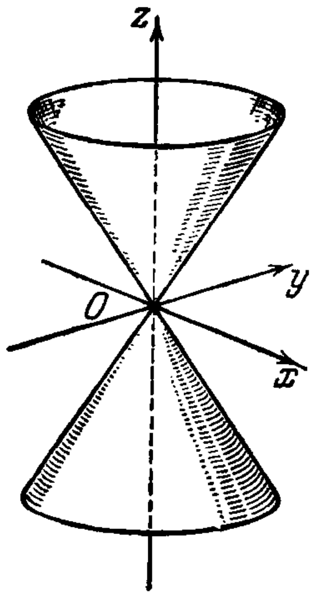


Рис. 57.

ходящей через ось Oz, чтобы найти ее «меридиан», т. е. линию, лежащую в плоскости, проходящей через ось, вращением которой она получается.

Пересечем поверхность (20) координатной плоскостью Oxz, т. е. плоскостью $y = 0$ (рис. 54), подставляя $y = 0$ в уравнение (20), и мы получим уравнение $\frac{x^2}{a^2} - \frac{z^2}{c^2} = \delta$ меридиана. В случае $\delta = 1$ — это гипербола I, при $\delta = -1$ — гипербола II, а при $\delta = 0$ — пара пересекающихся прямых III. При вращении они дают так называемый однополостный **гиперболоид вращения** (рис. 55), двуполостный гиперболоид вращения (рис. 56) и прямой круговой конус (рис. 57).

Общие однополостный гиперболоид, двуполостный гиперболоид и конус 2-го порядка (3, 4 и 5) получаются из только что рассмотренных поверхностей вращения растяжением их от плоскости Oxz с коэффициентом b/a.

Параболоиды. Остаются еще уравнения 7) и 8). Первое из них, $\frac{x^2}{a^2} + \frac{y^2}{b^2} = 2cz$, сравниваем с уравнением

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

которое исследуем аналогично рассмотренному в предыдущем пункте и убеждаемся, что оно выражает собою поверхность, получаемую от враще-

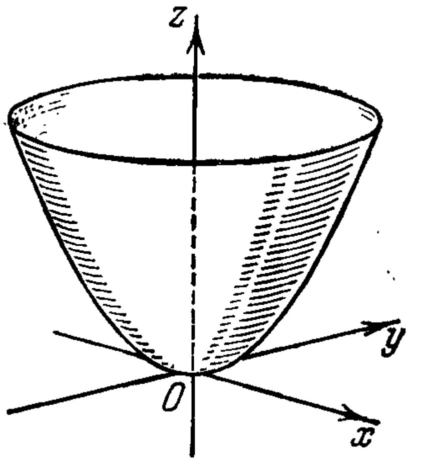


Рис. 58.

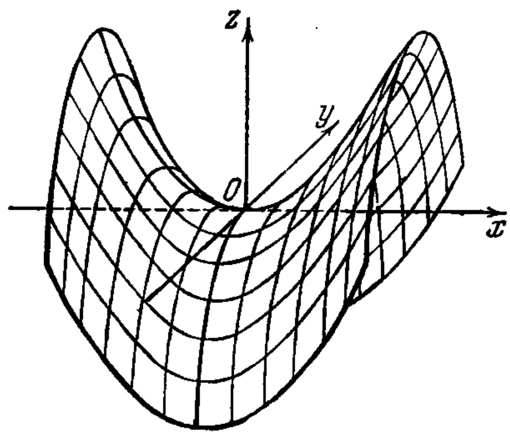


Рис. 59.

ния параболы $x^2 = 2a^2cz$ вокруг оси Oz,— так называемый параболоид вращения (рис. 58), о котором мы уже говорили, когда рассматривали параболические зеркала. Общий эллиптический параболоид 7) получается из параболоида вращения растяжением от плоскости Oxz.

Поверхность 8) приходится исследовать иначе, именно: исследуя ее сечения плоскостями $z = h$, которые суть гиперболы. Карта поверхности 8) в горизонталях изображена на рис. 50; при другом положении координатных осей мы рассматривали эту поверхность на рис. 51. Эта поверхность седлообразная. Она имеет вид, изображенный на рис. 59, и называется гиперболическим параболоидом. Сечения ее плоскостями, параллельными плоскости Oxz, оказываются одинаковыми параболами. То же самое дают сечения плоскостями, параллельными плоскости Oyz.

Прямолинейные образующие однополостного гиперболоида. Весьма любопытно, но совершенно не очевидно, что однополостный гиперболоид и гиперболический параболоид, подобно конической или цилиндрической поверхности, могут быть получены движением прямой линии. В случае гиперболоида, достаточно это доказать для однополостного гиперболоида вращения $\frac{x^2}{a^2}+\frac{y^2}{a^2}-\frac{z^2}{c^2}=1$, так как общий однополостный гиперболоид получается из него равномерным растяжением от плоскости Oxz, а при таком растяжении любая прямая снова переходит в прямую. Пересечем

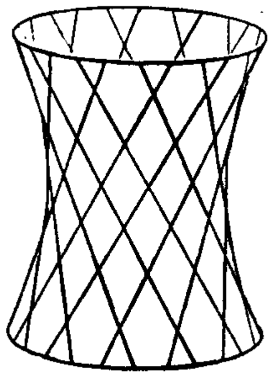


Рис. 60.

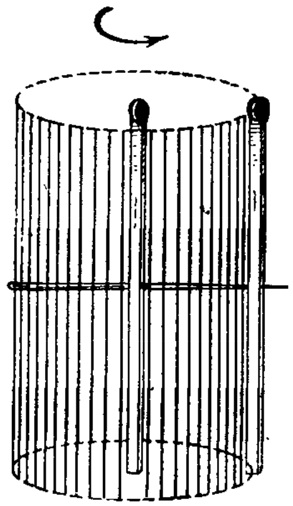


Рис. 61.

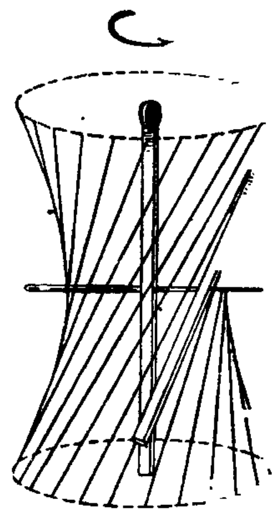


Рис. 62

гиперболоид вращения плоскостью $y=a$, параллельной плоскости Oxz. Подставляя $y=a$, получим

$$\frac{x^2}{a^2}+\frac{a^2}{a^2}-\frac{z^2}{c^2}=1 \quad \text{или} \quad \frac{x^2}{a^2}-\frac{z^2}{c^2}=0.$$

Но это уравнение вместе с $y=a$ дает в плоскости $y=a$ пару пересекающихся прямых: $\frac{x}{a}-\frac{z}{c}=0$ и $\frac{x}{a}+\frac{z}{c}=0$.

Итак, мы уже обнаружили, что на гиперболоиде лежит пара пересекающихся прямых. Если теперь вращать гиперболоид вокруг оси Oz, то каждая из этих прямых, очевидно, зачертит весь гиперболоид (рис. 60).

Легко доказать, что: 1) любые две прямые одного и того же из получающихся семейств не лежат в одной плоскости (т. е., как говорят, скрещиваются), 2) любая прямая одного из этих семейств пересекает все прямые другого (кроме ей противоположной, которая ей параллельна) и 3) три прямые одного и того же семейства не параллельны никакой одной и той же плоскости.

При помощи двух спичек и иголки легко получить представление о однополостном гиперболоиде вращения. Если проколоть иголкой одну спичку в ее середине насквозь и на выступающее острие иголки наколоть серединой другую спичку в параллельном к первой спичке положении, то при вращении вокруг первой спички, как вокруг оси, вторая спичка будет описывать поверхность цилиндра (рис. 61). Если же вторую спичку наколоть перпендикулярно иголке, но не параллельно первой спичке, то она при таком вращении будет описывать поверхность однополостного гиперболоида вращения, которую при быстром вращении хорошо видно (рис. 62).

Итог исследования уравнения 2-й степени. Хотя общее уравнение 2-й степени с тремя переменными может выражать 17 существенно различных поверхностей, запомнить их нетрудно. Девять последних из них суть цилиндры над девятью возможными линиями 2-го порядка. Восемь же первых разбиваются на четыре пары: два эллипсоида (действительный и мнимый), два гиперболоида (однополостный и двуполостный), два конуса 2-го порядка (действительный и мнимый) и два параболоида (эллиптический и гиперболический). Все эти поверхности играют существенную роль в механике, физике и технике. (Эллипсоид инерции, эллипсоид упругости, гиперболоид в преобразовании Лоренца в физике, параболоид вращения для параболических зеркал и т. д.).

§ 11. ПРЕОБРАЗОВАНИЯ АФФИННЫЕ И ОРТОГОНАЛЬНЫЕ

Следующим важным этапом развития аналитической геометрии было введение в нее, и вообще в геометрию, теории преобразований. Тут придется подробно пояснить в чем дело.

«Сжатие» плоскости к прямой. Рассмотрим одно из самых простых преобразований плоскости — равномерное «сжатие» к прямой с коэффициентом k. Пусть на плоскости дана прямая a и дан положительный коэффициент k, например $k = \frac{2}{3}$. Оставим все точки прямой a на месте, а всякую точку M, не лежащую на этой прямой, заменим точкой M', такой, что M' лежит по ту же сторону от прямой a, где и точка M, на том же перпендикуляре к a, на котором лежит точка M; но так, что расстояние от точки M' до прямой a равно $^2/_3$ расстояния до нее от точки M. Если коэффициент k, как здесь, меньше единицы, то происходит собственно сжатие плоскости к прямой; если же k больше единицы, то происходит растяжение плоскости от прямой, но для удобства мы будем и в том и в другом случае говорить о «сжатии», только слово «сжатие» будем ставить в кавычки.

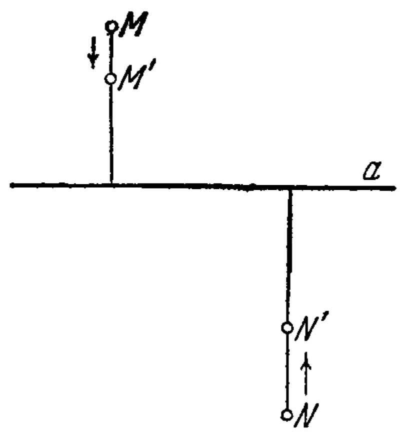


Рис. 63.

Преобразуемая точка или фигура называется прообразом, а та, в которую она перешла,— ее образом. Точка M', например, есть образ точки M (рис. 63).

Покажем, что при равномерном «сжатии» плоскости к прямой всякая прямая плоскости преобразуется в прямую. Действительно, пусть плоскость «сжимается» к лежащей на ней прямой a с коэффициентом «сжатия» k. Пусть b — какая угодно прямая плоскости, O — точка, в которой она пересекает прямую a, B — произвольная другая ее точка, а BA — перпендикуляр, опущенный из этой точки на прямую a (рис. 64). После «сжатия» точка B перейдет в некоторую точку B' на этом перпендикуляре, такую, что $B'A = k \cdot BA$. Поэтому тангенс угла $B'OA$ будет равен $\frac{AB'}{OA} = \frac{k \cdot AB}{OA}$,

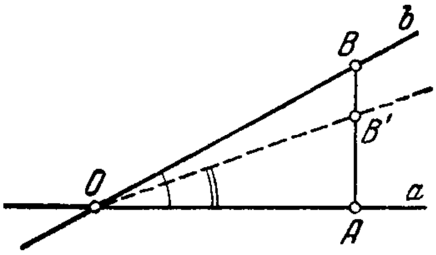


Рис. 64.

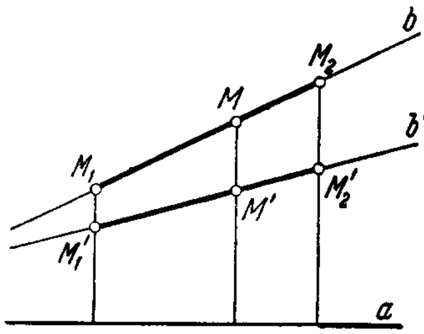


Рис. 65.

т. е. будет равен k раз взятому тангенсу угла, образуемого прямой b с прямой a, т. е. для всех точек B', в которые перейдут разные точки прямой b, он будет один и тот же. Все точки B' лежат, следовательно, на одной и той же прямой, проходящей через точку O и образующей с прямой a угол с таким тангенсом.

При «сжатии» параллельные прямые остаются параллельными. Действительно, если тангенсы углов, которые образуют прямые b и c с прямой a, одинаковы, то тангенсы тех углов, которые образуют с a их образы b' и c', отличаются от них только множителем k, т. е. между собою тоже одинаковы, т. е. прямые b' и c' также между собою параллельны.

Всякий прямолинейный отрезок плоскости при «сжатии» плоскости к прямой сокращается (или удлиняется) равномерно (хотя и в разной степени для отрезков разных направлений). Говоря о «равномерном» сокращении, мы подразумеваем, что середина отрезка остается серединой, треть — третью и т. д., т. е. отрезок сжимается равномерно по всей своей длине. Действительно, в каком отношении точка M делит отрезок M_1M_2, в таком же и ее образ M' делит образ $M_1'M_2'$ этого отрезка, так как параллельные прямые (у нас перпендикуляры к прямой a) разрезают секущие их прямые (в этом случае b и b') на пропорциональные части (рис. 65).

Эллипс как результат «сжатия» окружности. Рассмотрим окружность с центром в начале и радиусом a. В силу теоремы Пифагора ее уравнение есть $x^2 + \bar{y}^2 = a^2$. Мы пишем не y, а $\bar{y}$, так как y нам понадобится дальше. Посмотрим, во что эта окружность превратится, если сделать «сжатие» плоскости к оси Ox с коэффициентом $\frac{b}{a}$ (рис. 66). После такого «сжатия» значения x всех точек останутся прежними, а значения $\bar{y}$ станут равными $y = \bar{y}\frac{b}{a}$, т. е. $\bar{y} = \frac{a}{b}y$. Подставляя $\bar{y}$ в написанное уравнение окружности, мы будем иметь: $x^2 + \frac{a^2}{b^2}y^2 = a^2$

или $\frac{x^2}{a^2}+\frac{y^2}{b^2}=1$, — это будет в тех же осях уравнением той линии, которая получилась из рассматриваемой окружности сжатием к оси Ox. Как мы видим, получается эллипс. Итак, мы доказали, что эллипс есть результат «сжатия» окружности.

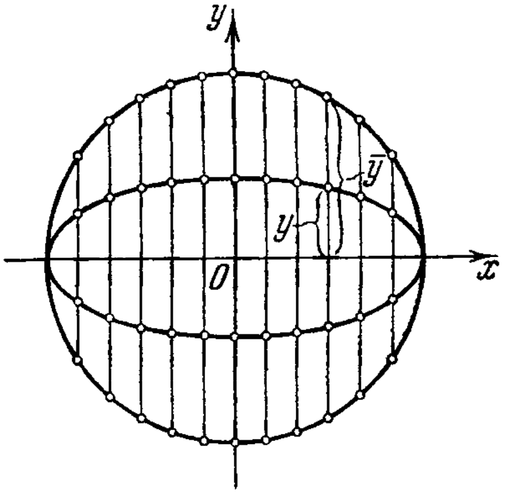


Рис. 66.

Из того, что эллипс есть «сжатие» окружности, непосредственно следуют многие свойства эллипса. Например, упомянутое свойство диаметров, что если даны параллельные секущие эллипса, то середины образуемых ими хорд лежат на прямой (см. рис. 12), может быть доказано следующим образом. Сделаем обратное растяжение эллипса в круг. При этом параллельные хорды эллипса превратятся в параллельные хорды круга, а их середины — в середины этих хорд. Но середины параллельных хорд круга лежат на его диаметре, т. е. на прямой, следовательно, и середины параллельных хорд эллипса также лежат на одной прямой. Именно, они лежат на той прямой, которая получается из диаметра круга тем «сжатием», которым из этой окружности получается эллипс.

А вот другое приложение теории «сжатия». Так как всякая вертикальная полоска круга при его сжатии к оси Ox, не изменяя своей ширины, меняет свою длину в b/a раз, то площадь этой полоски после сжатия равна ее исходной площади, помноженной на b/a, а так как площадь круга равна πa^2, то площадь соответственного эллипса равна $\pi a^2 \frac{b}{a} = \pi ab$.

Пример решения более сложной задачи. Пусть задан эллипс и надо найти треугольник наименьшей площади, описанный вокруг этого эллипса. Решим сначала задачу для окружности. Покажем, что в случае окружности — это равносторонний треугольник. Действительно, пусть описанный треугольник неравносторонний, т. е. наименьший его угол (обозначим его B) меньше $60°$, а наибольший $C>60°$. Тогда, не меняя угла A, передвинем сторону BC в положение B_0C_0 (рис. 67), перемещая вершину B к A до тех пор, пока один из углов B_0 или C_0 станет равным $60°$. Мы получим описанный треугольник AB_0C_0 с меньшей площадью, так как при этом[1] будет $OC<OB$, $OC_0\leqslant OB_0$, и потому отброшенная площадь OBB_0 больше введенной OCC_0. Если полученный треугольник неравносторонний, то, повторяя второй раз все рассуждения, мы еще уменьшим его площадь и придем к равностороннему. Значит, всякий неравносторонний треугольник, описанный вокруг данной окружности, имеет бо́льшую площадь, чем равносторонний.

[1] Как это легко показать.

Вернемся теперь к эллипсу. Сделаем такое растяжение его от его большой оси, которое превратит его обратно в окружность, из которой он получен «сжатием». При таком растяжении (рис. 68): 1) все описанные вокруг эллипса треугольники обратятся в треугольники, описанные во-

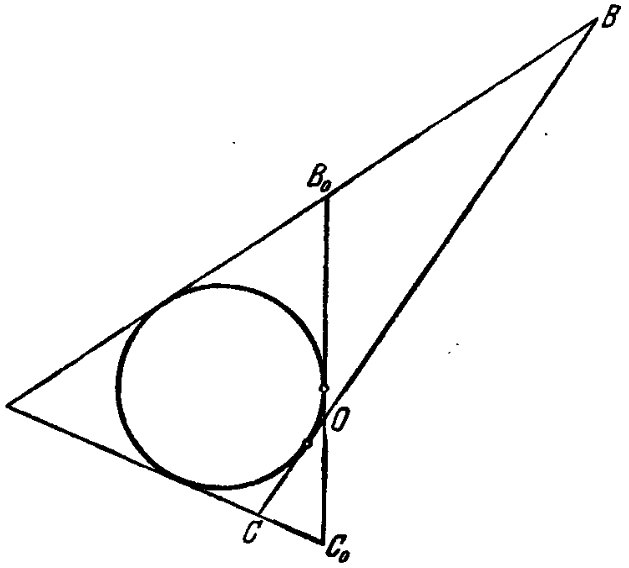


Рис. 67.

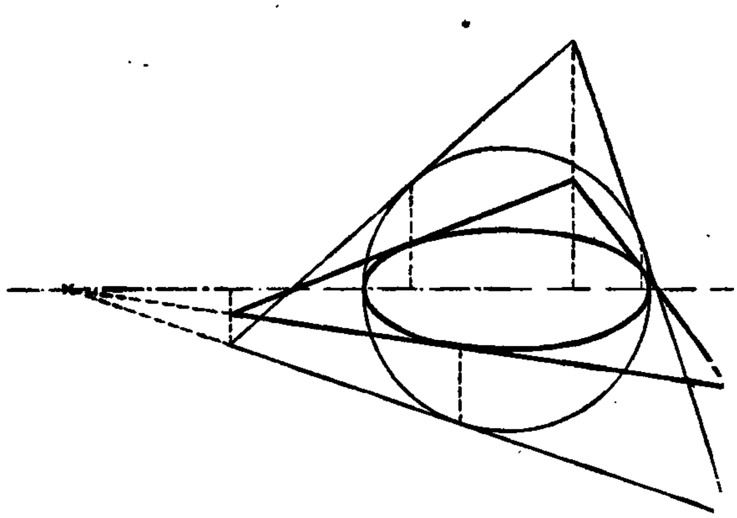


Рис. 68.

круг полученной окружности; 2) площади всех фигур, в частности и всех этих треугольников, увеличатся в одно и то же число раз. Отсюда мы видим, что наименьшими по площади треугольниками, описанными вокруг данного эллипса, будут те, которые превращаются в равносторонние треугольники, описанные вокруг окружности. Таких треугольников бесконечно много, их центры тяжести лежат в центре эллипса, точки касания будут в серединах их сторон. Всякий из них легко построить (рис. 68), исходя от упоминавшейся только что окружности.

«Сжатие» плоскости к прямой — это только частный случай более общих, так называемых *аффинных*, преобразований плоскости.

Общие аффинные преобразования. Мы будем называть координатным «репером» плоскости пару векторов e_1, e_2, исходящих из общего начала O и не лежащих на одной прямой. Координатами точки M плоскости по отношению к такому реперу Oe_1e_2 будут тогда числа x, y такие, что для того, чтобы дойти от начала O до точки M, надо отложить x раз от точки O вектор e_1 и затем y раз вектор e_2. Это — общие декартовы координаты на плоскости. Аналогично определяются общие декартовы координаты в пространстве. Обычные так называемые прямоугольные декартовы координаты, которыми мы пользовались до сих пор, соответствуют тому частному случаю, когда координатные векторы e_1, e_2 взаимно перпендикулярны, а длины их равны длине той масштабной единицы, которой мы измеряем все отрезки.

Общее аффинное преобразование плоскости — это такое, при котором заданная сетка равных параллелограмов превращается в произвольную другую сетку равных параллелограмов. Точнее говоря, это такое преобразование плоскости, при котором заданный координатный репер Oe_1e_2

преобразуется в некоторый другой репер (вообще говоря, с другой «метрикой», т. е. с другими длинами своих векторов e'_1 и e'_2 и другим углом между ними), а любая точка M преобразуется в точку M', имеющую те же координаты относительно нового репера, какие M имела относительно старого (рис. 69).

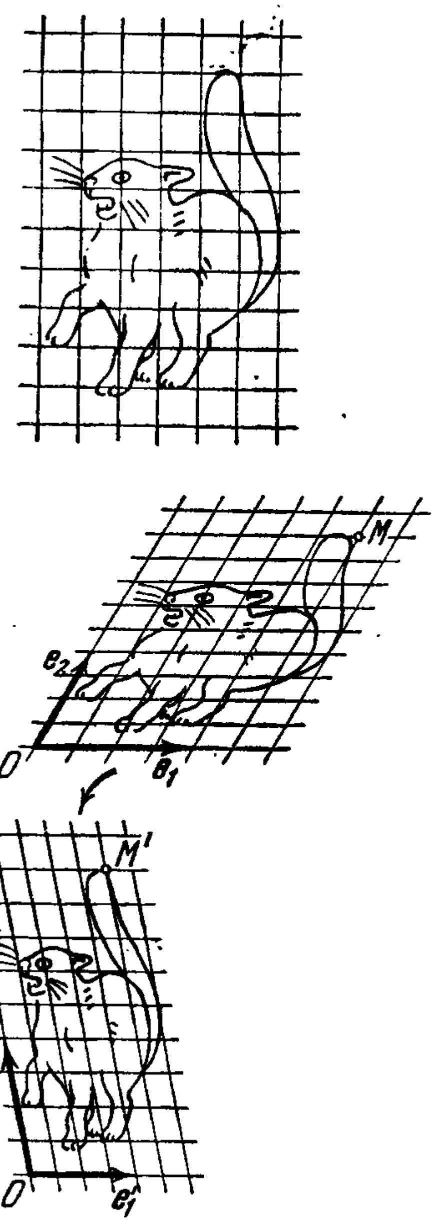


Рис. 69.

«Сжатие» к оси Ox с коэффициентом k есть такой частный случай, когда прямоугольный репер $Oe'_1e'_2$ переходит в репер Oe'_1ke_2.

Можно легко показать, что и при общем аффинном преобразовании любая прямая переходит в прямую, параллельные прямые переходят в параллельные, а если точка делит отрезок в некотором отношении, то ее образ делит образ этого отрезка в том же отношении. Кроме того, можно доказать замечательную теорему, что любое аффинное преобразование плоскости можно получить, сделав некоторое движение плоскости в себе, как жесткого целого, и затем два «сжатия», вообще говоря, с разными коэффициентами k_1 и k_2, к некоторым двум взаимно перпендикулярным прямым.

Для доказательства этого утверждения рассмотрим все радиусы некоторой окружности преобразуемой плоскости (рис. 70). Пусть радиус OA — тот, который после преобразования оказывается самым коротким, пусть он превращается в $O'A'$. Перпендикуляр AB к OA превращается тогда в перпендикуляр же $A'B'$ к $O'A'$, так как, если бы перпендикуляр $O'C'$ отличался от $O'A'$, то он был бы образом наклонной OC и тогда образ $O'D'$ радиуса OD был бы частью перпендикуляра $O'C'$, т. е. короче наклонной $O'A'$, вопреки предположению.

Взаимно перпендикулярные прямые OA и AB поэтому переходят во взаимно перпендикулярные $O'A'$ и $A'B'$. Следовательно, квадратная сетка, построенная на OA и AB, переходит в сетку равных прямоугольников (рис. 71), и происходят равномерные «сжатия» вдоль прямых этой квадратной сетки.

Совершенно аналогично определяется общее аффинное преобразование пространства как такое, при котором пространственный координатный репер $Oe_1e_2e_3$ преобразуется в некоторый другой репер $O'e'_1e'_2e'_3$, вообще говоря, с другой «метрикой», т. е. с другими длинами единичных отрезков и другими углами между ними, а точки M — в точки M', имеющие такие же координаты относительно нового репера, какие имели точки M относительно старого.

Все перечисленные нами свойства имеются и у аффинных преобразований пространства, только в последней теореме, в случае пространства

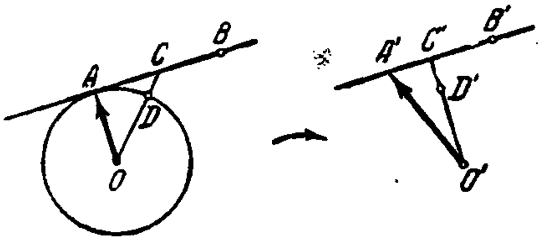


Рис. 70.

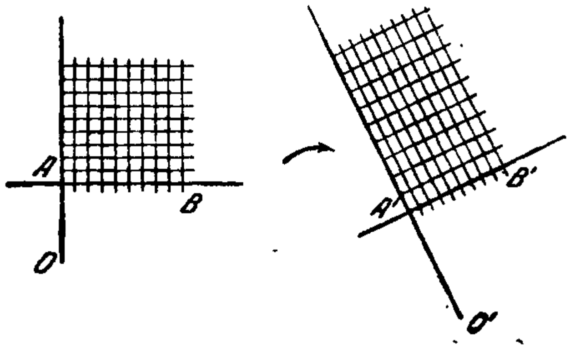


Рис. 71.

речь будет идти о движении пространства, как жесткого целого, и затем о трех «сжатиях» к трем взаимно перпендикулярным плоскостям с некоторыми коэффициентами k_1, k_2, k_3.

Важнейшие применения аффинных преобразований

1) Применение в геометрии для решения задач на аффинные свойства фигур, т. е. такие свойства, которые сохраняются при аффинных преобразованиях. Теорема о диаметрах эллипса и задача об описанном треугольнике были примерами таких задач. Для решения подобных задач аффинно преобразуют фигуру в какую-нибудь более простую, на ней обнаруживают искомое свойство и затем возвращаются к исходной фигуре.

2) Применение в аналитической геометрии для классификации линий и поверхностей 2-го порядка. Дело в том, что, как можно доказать, различные эллипсы родственны друг другу в том смысле, что получаются друг из друга аффинными преобразованиями (affinis собственно и значит по-латински родственный). Также и все гиперболы аффинны друг другу и все параболы аффинны друг другу. Но эллипс в гиперболу или параболу или гипербола в параболу уже не могут быть превращены ни при каком аффинном преобразовании, т. е. они аффинно друг другу не родственны. Естественно разбить все линии 2-го порядка на аффинные классы аффинно родственных друг другу линий. Оказывается, что приведение уравнения к каноническому виду как раз и дает эту классификацию, т. е. что аффинных классов линий 2-го порядка девять. (Мы не будем входить в подробности того, почему мнимые эллипсы и пары мнимых параллельных прямых относят к разным аффинным классам. Ни в том, ни в другом случае кривых на плоскости, собственно говоря, нет. Речь идет уже об алгебраических свойствах самого уравнения.)

Аналогично классификация поверхностей 2-го порядка по их каноническим уравнениям на 17 видов есть также аффинная классификация.

Дадим несложный пример применения аффинной классификации поверхностей 2-го порядка. Покажем, что если произвольно выбрать в пространстве три прямые a, b, c, такие, что 1) любые две из них не лежат в одной плоскости (т. е. скрещиваются) и 2) все три не параллельны одновременно одной и той же плоскости, то совокупность всех прямых d пространства, из которых каждая одновременно пересекает все три рассматриваемые прямые a, b и c (рис. 72), образует полную поверхность некоторого однополостного гиперболоида.

Поясним, о какой совокупности прямых d мы здесь говорим. Через любую точку A прямой a можно провести плоскость P, содержащую прямую b, и плоскость Q, содержащую прямую c. Эти плоскости P и Q пересекутся по той единственной прямой d, которая проходит через точку A

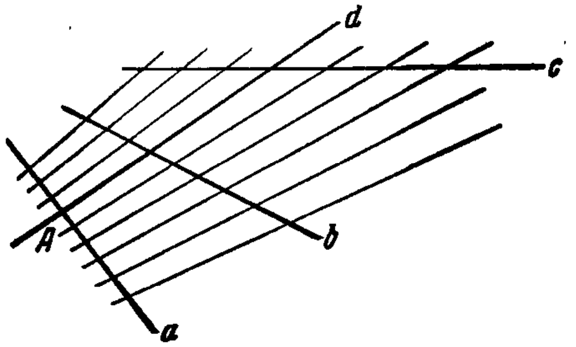


Рис. 72.

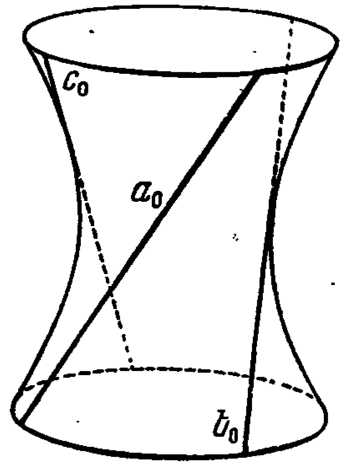


Рис. 73.

прямой a и пересекает прямые b и c. Проводя такие прямые d через любую точку прямой a, мы получим совокупность всех прямых d пространства, из которых каждая пересекает все три заданные прямые a, b и c. Эта совокупность прямых образует некоторую поверхность. Заметим, что всякий заданный однополостный гиперболоид можно получить таким способом, стоит лишь за прямые a, b и c взять три разные прямолинейные образующие a_0, b_0, c_0 одного семейства (рис. 73), а за прямые d — все образующие другого семейства. Пусть, наоборот, даны три какие угодно попарно скрещивающиеся прямые пространства a, b, c, одновременно не параллельные одной и той же плоскости. Тогда, как можно показать, они всегда суть прямые трех попарно не имеющих общих точек ребер некоторого параллелепипеда (рис. 74). Построив так параллелепипеды для заданных прямых a, b, c и для трех образующих a_0, b_0 c_0 одного и того же семейства какого угодно однополостного гиперболоида, сделаем то аффинное преобразование пространства, которое преобразует параллелепипед a_0, b_0, c_0 в параллелепипед a, b, c; оно, очевидно, преобразует этот гиперболоид в рассматриваемую поверхность. Но в силу аффинной классификации поверхностей

2-го порядка аффинный образ однополостного гиперболоида есть снова однополостный гиперболоид.

3) Применение к теории непрерывных преобразований сплошной среды, например в теории упругости, в теории течения жидкости, в теории электрического или магнитного поля и т. д. Очень малые элементы рассматриваемой сплошной среды преобразуются «почти» аффинно. Как говорят, «в малом преобразование линейно» (линейными называются выражения 1-й степени, а в следующем пункте мы увидим, что в аналитической геометрии формулы аффинных преобразований именно 1-й степени).

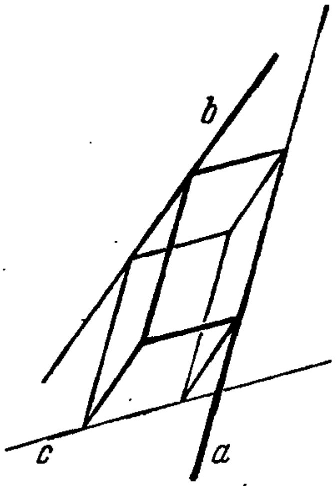


Рис. 74.

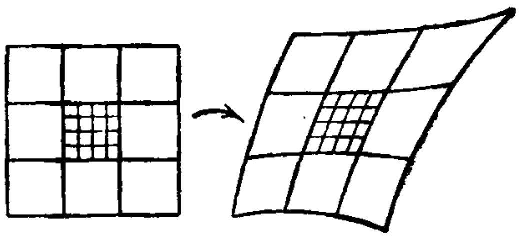


Рис. 75.

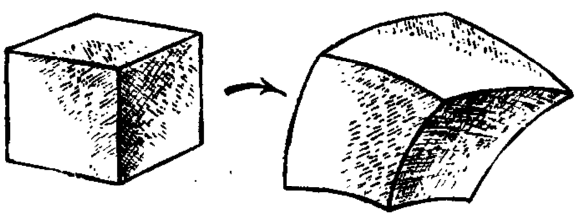


Рис. 76.

Это видно на рис. 75. На прямых крупной квадратной сетки ясно заметно их искривление, расхождение «веером» и т. д. Для небольшого же кусочка весьма густой квадратной сетки все это уже весьма мало сказывается, и он преобразуется «почти» в сетку равных параллелограмов. Аналогичная картина получается и в пространстве (рис. 76). В силу того, что всякое аффинное преобразование пространства сводится к движению и трем взаимно перпендикулярным «сжатиям», отсюда следует, что элемент тела при упругой его деформации, во-первых, передвигается, как жесткое тело, и, кроме того, подвергается трем взаимно перпендикулярным «сжатиям».

Формулы аффинных преобразований. Если Oe_1e_2 — аффинно преобразуемый репер, $O'e_1'e_2'$ — его образ, и координаты нового начала O' относительно старого репера суть ξ, η, а координаты векторов e_1' и e_2' относительно него a_1, a_2 и b_1, b_2, то формулы аффинного преобразования, как легко видеть из рис. 77, суть

$$x' = a_1x + b_1y + \xi,$$
$$y' = a_2x + b_2y + \eta$$

в том смысле, что если x, y — координаты любой точки M относительно старого репера Oe_1e_2, то x', y', выражаемые этими формулами, суть координаты по отношению к тому же старому реперу образа M' этой точки.

Действительно, пусть Oe_1e_2 — преобразуемый репер, $O'e_1'e_2'$ — его образ, M — какая-нибудь преобразуемая точка плоскости, а M' — ее образ. Тогда в силу самого определения аффинного преобразования, если координаты точки M относительно репера Oe_1e_2 суть x, y, то координаты ее образа M' по отношению к образу $O'e_1'e_2'$ этого репера — те же самые x, y.

Рассмотрим теперь вектор m', идущий из начала O преобразуемого репера в образ M' точки M. Тогда $m' = x'e_1 + y'e_2$. Но этот вектор равен векторной сумме

$$m' = \xi e_1 + \eta e_2 + xe_1' + ye_2',$$

а векторы e_1' и e_2' суть

$$e_1' = a_1e_1 + a_2e_2, \quad e_2' = b_1e_1 + b_2e_2$$

и, следовательно,

$$m' = \xi e_1 + \eta e_2 + a_1xe_1 + a_2xe_2 + b_1ye_1 + b_2ye_2$$

или

$$m' = (a_1x + b_1y + \xi)\, e_1 + (a_2x + b_2y + \eta)\, e_2.$$

Сравнивая полученное выражение с первым выражением для m', получаем

$$\left.\begin{aligned} x' &= a_1x + b_1y + \xi,\\ y' &= a_2x + b_2y + \eta.\end{aligned}\right\} \qquad (21)$$

Определитель

$$\Delta = \begin{vmatrix} a_1 & b_1 \\ a_2 & b_2 \end{vmatrix} = a_1b_2 - a_2b_1,$$

как можно доказать, будет не равен нулю и равен отношению площади параллелограма, построенного на векторах нового репера, к площади такого же параллелограма, построенного на векторах старого.

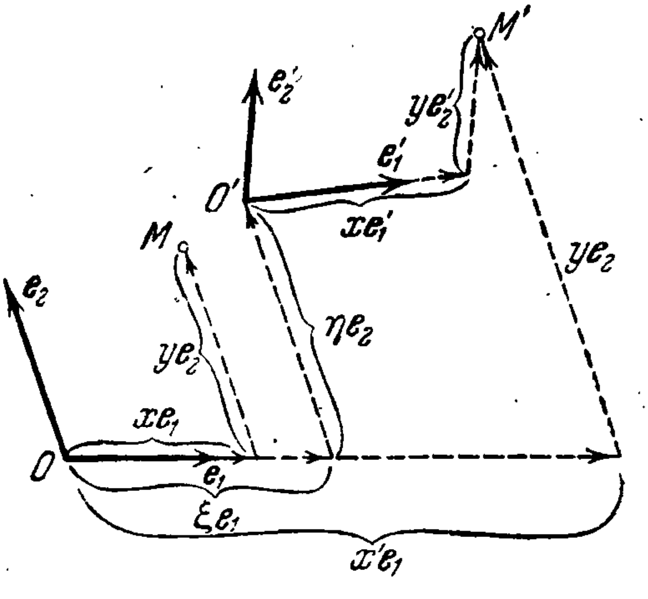


Рис. 77.

Аналогичные формулы получаются для пространства

$$\left.\begin{aligned} x' &= a_1x + b_1y + c_1z + \xi,\\ y' &= a_2x + b_2y + c_2z + \eta,\\ z' &= a_3x + b_3y + c_3z + \zeta;\end{aligned}\right\} \qquad (22)$$

где (ξ, η, ζ) — координаты начала O' преобразованного репера $O'e_1'e_2'e_3'$, а (a_1, a_2, a_3), (b_1, b_2, b_3), (c_1, c_2, c_3) — координаты его векторов e_1', e_2', e_3 относительно преобразуемого репера $Oe_1e_2e_3$.

Определитель [1]

$$\Delta = \begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix} = a_1b_2c_3 + a_2b_3c_1 + a_3b_1c_2 - a_1b_3c_2 - a_2b_1c_3 - a_3b_2c_1$$

не равен нулю и равен отношению объема параллелепипеда, построенного

[1] Об определителях см. главу XVI (том 3).

на векторах нового репера, к объему параллелепипеда, построенного на векторах исходного репера.

Ортогональные преобразования. Движения плоскости по себе, как жесткого целого, или такие движения плюс отражение в некоторой прямой, лежащей на плоскости, называются *ортогональными преобразованиями плоскости*, а движения пространства, как жесткого целого, или такие движения плюс отражение пространства в некоторой его плоскости называются *ортогональными преобразованиями пространства*. Очевидно, что ортогональные преобразованин — это такие аффинные преобразования, при которых не изменяется «метрика» репера, а он только претерпевает передвижение, либо передвижение плюс отражение.

Будем исследовать ортогональное преобразование при помощи прямоугольных координат, т. е. когда векторы исходного репера взаимно перпендикулярны и имеют длины, равные абсолютной масштабной единице. После ортогонального преобразования векторы репера останутся взаимно перпендикулярными, т. е. их скалярные произведения будут равны нулю и останутся по длине равными единице. Поэтому (см. формулу (14) на стр. 210) в случае плоскости мы будем иметь

$$a_1b_1 + a_2b_2 = 0,\ a_1^2 + a_2^2 = 1,\ b_1^2 + b_2^2 = 1, \tag{21'}$$

а в случае пространства

$$\begin{aligned} a_1b_1 + a_2b_2 + a_3b_3 &= 0,\ a_1^2 + a_2^2 + a_3^2 = 1, \\ a_1c_1 + a_2c_2 + a_3c_3 &= 0,\ b_1^2 + b_2^2 + b_3^2 = 1, \\ b_1c_1 + b_2c_2 + b_3c_3 &= 0,\ c_1^2 + c_2^2 + c_3^2 = 1. \end{aligned} \tag{22'}$$

Поэтому, если исходный репер считать прямоугольным, то формулы (21) дают ортогональное преобразование плоскости тогда и только тогда, когда выполняются условия ортогональности (21′), а формулы (22) дают ортогональное преобразование пространства, если выполняются условия ортогональности (22′). Можно показать, что если $\Delta > 0$, то это — движение, если $\Delta < 0$, то это — движение плюс отражение.

§ 12. ТЕОРИЯ ИНВАРИАНТОВ[1]

Идея инварианта. Инварианты уравнения 2-й степени с двумя переменными. Во второй половине прошлого века было введено еще одно важное новое понятие — понятие инварианта.

Рассмотрим, например, многочлен 2-й степени с двумя переменными

$$Ax^2 + 2Bxy + Cy^2 + 2Dx + 2Ey + F. \tag{23}$$

Если рассматривать x, y как прямоугольные координаты и производить преобразование их к новым прямоугольным осям, то после подстановки в (23) вместо x, y их выражений через новые координаты x', y', раскрытия скобок и приведения подобных членов мы получим новый преобразованный многочлен с другими коэффициентами

$$A'x'^2 + 2B'x'y' + C'y'^2 + 2D'x' + 2E'y' + F'. \tag{24}$$

---

[1] Invarians по-латински неизменный.

Оказывается, существуют такие выражения, составленные из коэффициентов, которые при этом преобразовании численно не меняются, хотя сами коэффициенты и меняются. Такое выражение от A', B', C', D', E', F' имеет ту же численную величину, какую оно имело бы, если бы его составить из A, B, C, D, E, F.

Выражения такого рода называются *инвариантами* многочлена (23) по отношению к группе ортогональных преобразований (т. е. по отношению к преобразованиям от одних прямоугольных координат x, y к любым другим прямоугольным координатам x', y').

Такими инвариантами, оказывается, будут

$$I_1 = A + C,$$

$$I_2 = \begin{vmatrix} A & B \\ B & C \end{vmatrix} = AC - B^2,$$

$$I_3 = \begin{vmatrix} A & B & D \\ B & C & E \\ D & E & F \end{vmatrix} = ACF + 2BDE - AE^2 - CD^2 - FB^2,$$

т. е.

$$A + C = A' + C', \quad AC - B^2 = A'C' - B'^2,$$

$$ACF + 2BDE - AE^2 - CD^2 - FB^2 = \\ = A'C'F' + 2B'D'E' - A'E'^2 - C'D'^2 - F'B'^2.$$

Можно доказать важную теорему, что любой ортогональный инвариант многочлена (23) выражается через эти три основных инварианта.

Если мы приравняем многочлен (23) нулю, то получим уравнение некоторой линии 2-го порядка. Всякая величина, связанная с самой этой линией, а не с ее расположением на плоскости, очевидно, не будет зависеть от того, в каких координатах написано ее уравнение, и поэтому, если она выражается как-нибудь через коэффициенты, это выражение будет ортогональным инвариантом многочлена (23) и, следовательно, в силу высказанной теоремы будет выражаться через эти три основных инварианта. Сверх того, так как при умножении на любое заданное число t, отличное от нуля, всех шести коэффициентов рассматриваемого уравнения, линия, им выражаемая, остается прежней, то выражение через I_1, I_2, I_3 всякой величины, связанной с самой линией, должно быть непременно таким, чтобы при умножении в нем A, B, C, D, E, F на t число t в нем сокращалось. Рассматриваемое выражение, как говорят, должно быть однородным нулевой степени по отношению к A, B, C, D, E, F.

Проверим это на примере. Пусть, например, уравнение

$$Ax^2 + 2Bxy + Cy^2 + 2Dx + 2Ey + F = 0$$

выражает некоторый эллипс. Так как это уравнение вполне задает этот эллипс, то при помощи него, т. е. при помощи его коэффициентов, можно вычислить все основные величины, связанные с этим эллипсом. Например, можно вычислить его полуоси a и b, т. е. можно выразить полуоси через его коэффициенты. Выражения эти будут инвариантами и, следовательно, будут как-то выражаться через I_1, I_2, I_3. Путем приведения уравнения к каноническому виду и некоторых дальнейших вычислений действительно получается следующее (довольно сложное) выражение полуосей через I_1, I_2, I_3:

$$\sqrt{\frac{2|I_3|}{|I_2||I_1 \pm \sqrt{I_1^2 - 4I_2}|}},$$

причем выражение это однородно относительно A, B, C, D, E, F.

Из сказанного видно, что сами инварианты I_1, I_2, I_3, как хотя и однородные, но не нулевой степени, выражения от коэффициентов, не имеют прямого геометрического смысла — это алгебраические объекты.

Можно показать, что выражение

$$K_1 = \begin{vmatrix} A & D \\ D & F \end{vmatrix} + \begin{vmatrix} C & E \\ E & F \end{vmatrix} = AF - D^2 + CF - E^2$$

хотя и может изменяться при параллельных переносах, но не изменяется при чистых поворотах заданных прямоугольных осей, — это так называемый *полуинвариант*.

Чтобы дать пример применения инвариантов и полуинвариантов, приведем следующую таблицу, которая, если вычислить I_1, I_2, I_3 и K_1, сразу дает возможность определить по уравнению аффинный класс линии 2-го порядка, им выражаемой:

Признак класса	Название	Приведенное уравнение	Каноническое уравнение
$I_2 > 0,\ I_1 I_3 < 0$	эллипс		$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$
$I_2 > 0,\ I_1 I_3 > 0$	мнимый эллипс		$\frac{x^2}{a^2} + \frac{y^2}{b^2} = -1$
$I_2 > 0,\ I_3 = 0$	точка	$\lambda_1 x^2 + \lambda_2 y^2 + \frac{I_3}{I_2} = 0$	$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 0$
$I_2 < 0,\ I_3 \neq 0$	гипербола		$\frac{x^2}{a^2} - \frac{y^2}{b^2} = 1$
$I_2 < 0,\ I_3 = 0$	пара пересекающихся прямых		$\frac{x^2}{a^2} - \frac{y^2}{b^2} = 0$
$I_2 = 0,\ I_3 \neq 0$	парабола	$I_1 x^2 + 2\sqrt{-\frac{I_3}{I_1}}\, y = 0$	$x^2 = 2py$
$I_2 = 0,\ I_3 = 0,\ K_1 < 0$	пара параллельных прямых		$x^2 = a^2$
$I_2 = 0,\ I_3 = 0,\ K_1 > 0$	пара мнимых параллельных прямых	$I_1 x^2 + \frac{K_1}{I_1} = 0$	$x^2 = -a^2$
$I_2 = 0,\ I_3 = 0,\ K_1 = 0$	пара совпадающих прямых		$x^2 = 0$

В этой таблице выписаны условия, необходимые и достаточные для того, чтобы уравнение линии 2-го порядка приводилось к тому или иному из девяти канонических видов ($I_1 I_3$ обозначает произведение I_1 и I_3).

Пусть, например, дано уравнение $x^2 - 6xy + 5y^2 - 2x + 4y + 3 = 0$. Мы имеем $A = 1$, $B = -3$, $C = 5$, $D = -1$, $E = 2$, $F = 3$, откуда $I_1 = 6$, $I_2 = -4$, $I_3 = -9$. Выполнены условия 4-й строки таблицы: $I_2 < 0$, $I_3 \neq 0$, т. е. это — гипербола. Полуоси ее равны

$$\sqrt{\frac{2 \cdot 9}{4 \cdot |6 \pm \sqrt{36+16}|}} \approx 0{,}57 \text{ и } 1{,}93.$$

Коэффициенты «приведенных» уравнений (I), (II) и (III) выражаются через инварианты и полуинварианты так:

$$\lambda_1 x''^2 + \lambda_2 y''^2 + \frac{I_3}{I_2} = 0, \tag{I}$$

$$I_1 x''^2 + 2\sqrt{-\frac{I_3}{I_1}}\, y'' = 0, \tag{II}$$

$$I_1 x''^2 + \frac{K_1}{I_1} = 0, \tag{III}$$

где λ_1 и λ_2 — корни так называемого характеристического квадратного уравнения

$$\lambda^2 - I_1 \lambda + I_2 = 0.$$

Формулы (I—III) позволяют сразу вычислять полуоси a и b эллипса и гиперболы, параметр p параболы и расстояние $2a$ между параллельными прямыми. Формулы для полуосей были выписаны выше. Параметр p получается равным:

$$p = \sqrt{-\frac{I_3}{I_1^3}}, \quad \text{расстояние } 2a = 2\sqrt{-\frac{K_1}{I_1^2}}.$$

Совершенно аналогичная теория инвариантов и полуинвариантов, соответствующая таблица для определения аффинного класса и формулы коэффициентов приведенных уравнений могут быть выведены для поверхностей 2-го порядка в трехмерном пространстве.

Надо сказать, что все вышеизложенное освещает только смысл и значение тех инвариантов, которые рассматриваются в аналитической геометрии для линий и поверхностей 2-го порядка. Само понятие инварианта имеет, однако, несравненно более широкое значение.

Инвариантом некоторого изучаемого объекта по отношению к некоторым рассматриваемым его преобразованиям называется всякая величина (численная, векторная и т. п.), связанная с этим объектом, которая не изменяется при этих преобразованиях. В рассматриваемом вопросе объект — многочлен 2-й степени с двумя переменными (т. е. собственно его коэффициенты), преобразования — преобразования многочлена, получаемые при переходе от одних прямоугольных координат к другим.

Другой пример. Объект — данная масса данного газа при данной температуре. Преобразования — изменение объема или давления этой массы газа. Инвариант, по закону Бойля-Мариотта, произведение объема на давление. Можно говорить о длинах отрезков в пространстве или величинах углов, как инвариантах группы движений пространства, об отношениях, в которых точки делят отрезки, или об отношениях площадей, как об инвариантах группы аффинных преобразований пространства, и т. д.

Особенно важное значение имеют различные инварианты в физике.

§ 13. ПРОЕКТИВНАЯ ГЕОМЕТРИЯ

Перспективные проекции. Уже давно художники начали изучать законы перспективы. Это было нужно потому, что человек видит предметы в перспективной проекции на сетчатку глаза, при этом форма и взаим-

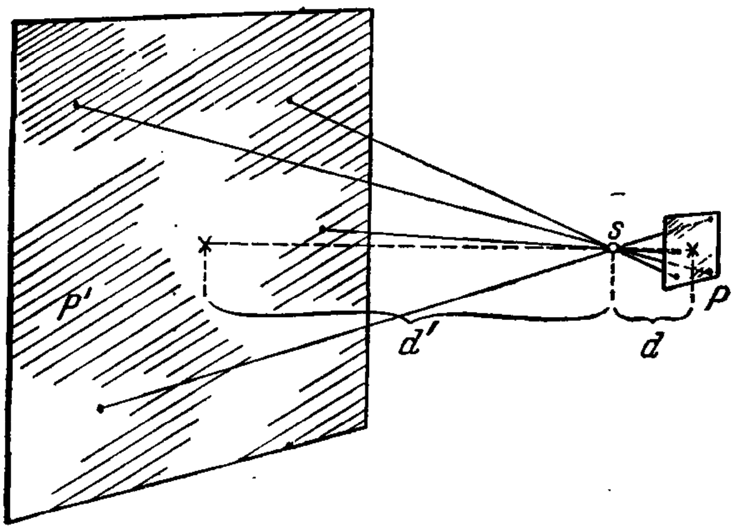


Рис. 78.

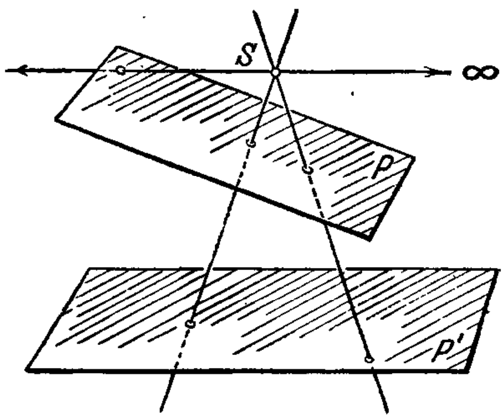


Рис. 79.

ное расположение предметов своеобразно искажаются. Например, уходящие вдаль телеграфные столбы как будто уменьшаются и сгущаются, параллельные рельсы железнодорожного пути кажутся сходящимися и т. д. Мы не будем здесь рассматривать пространственной перспективы, т. е. свойств перспективных проекций пространственных предметов на плоскость, а рассмотрим только свойства перспективных проекций плоскости на плоскость.

Пусть имеется снимок (например, кадр кинофильма) P, экран P' и расположенный между ними объектив S (рис. 78). Тогда, если снимок прозрачный и его осветить сзади (если он непрозрачный, то спереди, т. е. с той стороны, где находится объектив), то освещенные точки снимка излучают пучки света, собираемые объективом так, что они снова в виде точек проектируются на экран P'. Будем считать, что дело происходит так, как будто бы точки снимка P проектировались на экран P' прямыми, проходящими через оптический центр S объектива.

Все будет совсем просто, если плоскости P и P' параллельны. В таком случае на плоскости P', очевидно, получается подобное изображение всего того, что есть на плоскости P. Изображение это будет уменьшенное или увеличенное в зависимости от того, будет ли отношение $d' : d$, где d и d' — расстояния от центра объектива до плоскостей P и P', меньше или больше единицы.

Гораздо сложнее будет обстоять дело, если плоскости P и P' не параллельны (рис. 79). В этом случае при проектировании через точку S не только изменяется величина фигур, но и форма их искажается. Параллельные прямые могут при этом проектироваться как сходящиеся; отношение, в котором точка делит отрезок, может измениться и т. д. Вообще некоторые соотношения, которые остаются неизменными даже при любом аффинном отображении, здесь могут изменяться.

Такого рода проекция имеет место, например, при аэрофотосъемке. Самолет при полете качается, и поэтому жестко прикрепленный к нему фотоаппарат (рис. 80 а), вообще говоря, не обращен объективом строго вертикально вниз и в момент снимка большей частью расположен как-то косо, т. е. получается искаженное изображение местности (которую мы предполагаем плоскою).

Как выправить это изображение? Для этого надо изучить свойство проекций плоскости P на другую плоскость Π, вообще говоря, ей не параллельную,— прямыми, проходящими через точку S, не лежащую ни в плоскости P, ни в плоскости Π. Такие проекции называются *перспективными проекциями*.

Мы докажем дальше следующую замечательную теорему.

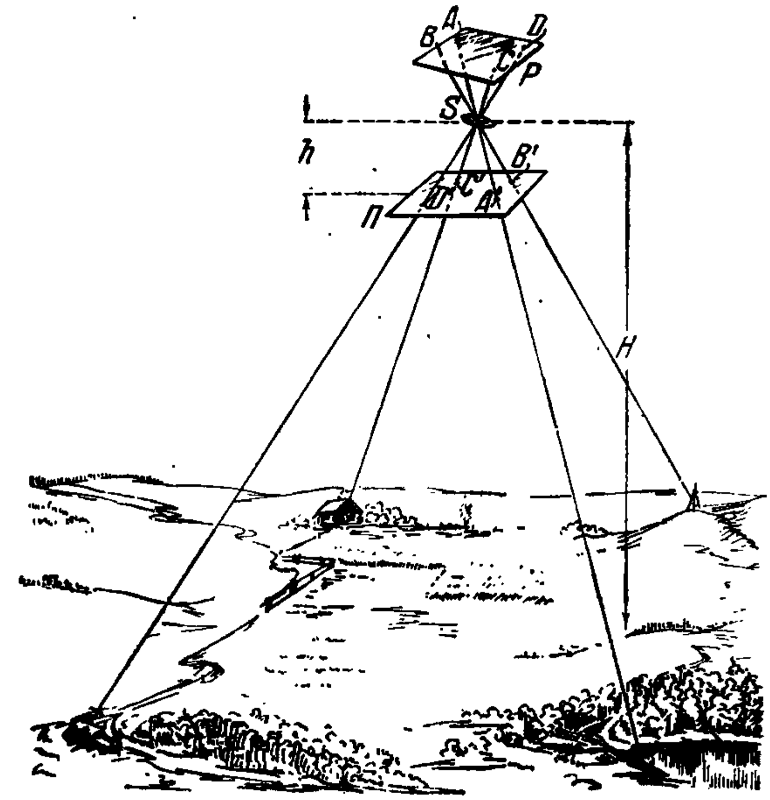
Теорема. Если имеются две перспективные проекции плоскости P на плоскость Π, такие, что при обеих проекциях точки A, B, C, D, образующие четверку точек «общего положения» плоскости P (т. е. такую четверку, у которой никакие три из ее точек не лежат на одной прямой), проектируются в одни и те же точки A', B', C', D' плоскости Π, то и все точки плоскости P проектируются при обеих проекциях в одни и те же точки плоскости Π.

Другими словами, результат перспективной проекции вполне определен, если известно, в какие точки при ней переходят точки какой-нибудь четверки точек общего положения проектируемой фигуры.

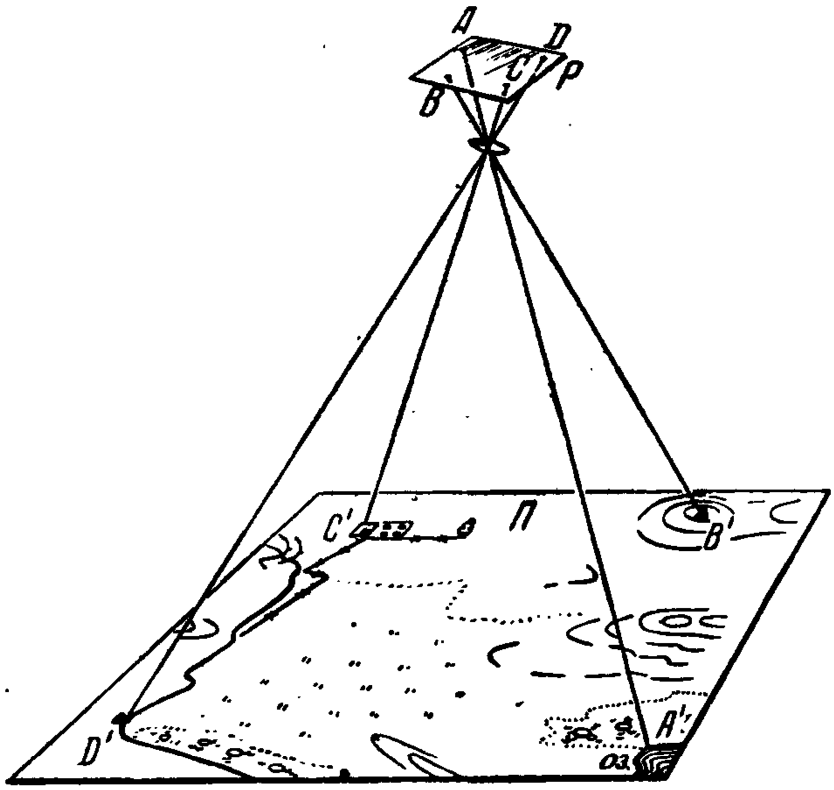
Это — так называемая теорема единственности теории проективных отображений или основная теорема плоской перспективы.

Применение основной теоремы плоской перспективы в аэрофотосъемке. Покажем, что эта теорема дает удобный способ выправить изображение при фотосъемке.

Если мысленно расположить в момент аэрофотоснимка горизонтальный экран Π так, чтобы высота центра S объектива над ним была h (рис. 80 а), то проекция на него через центр S изображения, снятого на фотопластинке P, будет, очевидно, не искаженной, а подобной картой



а.



б.

Рис. 80

горизонтальной местности с масштабом $h : H$, где H — высота самолета над местностью в момент снимка. Для того чтобы выправить изображение, полученное на снимке P, и превратить его в правильное, поступают так. Проявленный фотоснимок P вставляют в проекционный аппарат, укрепленный на специальном штативе, на котором этот аппарат с помощью установочных винтов можно удалять и приближать к экрану Π и всеми способами его поворачивать.

К экрану Π (рис. 80 б) прикрепляют топографическую карту местности, заснятую измерениями на земле (не подробную, тогда как фотоснимок дает все интересующие нас подробности, которых на карте может не быть). На этой прикрепленной к экрану Π карте выбирают четыре точки A', B', C', D', которые легко найти и на снимке (например, такой-то угол пересечения дорог, такой-то угол дома и т. д.), и в соответствующих точках A, B, C, D прокалывают иголкой снимок P, который производится на пленке. Сзади пластинки P вставляется проекционная лампа, так что снимок проектируется на экран Π через объектив S того аппарата, в который он вставлен. Оперируя установочными винтами, добиваются того, чтобы зайчики от проколов оказались в соответственных точках A', B', C', D' карты, прикрепленной к экрану Π. После того как это достигнуто, вместо карты вставляют кассету с фотопластинкой и, не сбивая установки, получают на фотопластинке снимок проекции на экран Π сделанного с самолета снимка P.

В силу высказанной выше теоремы на снимке получается правильная т. е. подобная местности), а не искаженная, карта снятого участка.

Перейдем теперь к изложению теории, которая нужна для доказательства указанной основной теоремы.

Проективная плоскость. Совокупность *всех* прямых и плоскостей пространства, проходящих через заданную точку пространства S, называется проектирующей связкой прямых и плоскостей с центром S. Если связку пересечь плоскостью P, не проходящей через ее центр, то всякой точке плоскости P будет соответствовать прямая связки, которая пересекает в этой точке плоскость P, а всякой прямой плоскости P — та плоскость связки, которая пересекает плоскость P по этой прямой. Однако так не устанавливается взаимно однозначное отображение совокупностей прямых и плоскостей связки на совокупности точек и прямых плоскости P. Дело в том, что прямым и плоскости связки, которые параллельны плоскости P, в этом смысле не соответствует никаких точек и никакой прямой плоскости P, так как они ее не пересекают. *Условливаются* всё же говорить, что эти прямые связки пересекают плоскость P, но в несобственных (или бесконечно удаленных) ее точках, лежащих в таких-то направлениях, и что эта плоскость связки пересекает плоскость P, но по несобственной (или бесконечно удаленной) ее прямой. Плоскость P, дополненная этими несобственными точками и несобственной прямой, называется *дополненной* или *проективной плоскостью*. Мы будем ее обозначать P^*. Совокупности прямых и плоскостей связки S уже взаимно однозначно отображаются на совокупности точек (собственных и несобственных) и прямых (собственных и несобственных) этой проективной плоскости P^*.

Кроме того, *уславливаются* говорить, что точка (собственная или несобственная) лежит на прямой (собственной или несобственной) проективной плоскости P^*, если соответственная прямая связки лежит в соответственной плоскости связки. С этой точки зрения любые две прямые проективной плоскости пересекаются (в собственной или несобственной точке), так как любые две плоскости связки пересекаются по некоторой прямой связки. Отсюда следует, между прочим, что несобственная прямая есть не что иное, как совокупность всех несобственных точек.

По сути дела, дополнение плоскости ее несобственными элементами сводится к тому, чтобы при помощи этой плоскости как сечения изучать связку всех прямых и плоскостей, проходящих через одну точку.

Проективные отображения; основная теорема. *Проективным* называется такое отображение проективной плоскости P^* на некоторую другую проективную плоскость $P^{*\prime}$ (которая может и совпадать с плоскостью P^*, и тогда говорят о проективном преобразовании плоскости P^*), которое, во-первых, является точечным взаимно однозначным отображением и, во-вторых, прямолинейно расположенные совокупности точек плоскости P^* переводит в прямолинейно же расположенные совокупности точек плоскости $P^{*\prime}$, и обратно. (При этом под точками и прямыми все время понимаются как собственные, так и несобственные точки и прямые.)

Очевидно, что две какие угодно перспективные проекции одной и той же плоскости P^* на некоторую плоскость Π^* получаются проективными преобразованиями друг из друга.

Действительно: 1° точки их (собственные и несобственные) взаимно однозначно связаны с точками (собственными и несобственными) проектируемой плоскости P^*, а следовательно, и между собою, и 2° прямолинейно расположенным точкам 1-й проекции соответствуют прямолинейно расположенные точки плоскости P^*, а, следовательно, и 2-й проекции, и обратно. Поэтому упомянутая выше теорема теории перспективы является непосредственным следствием такой теоремы о проективных преобразованиях: если при некотором проективном преобразовании плоскости Π^* точки ее A, B, C, D, образующие четверку общего положения, остаются на месте, то и все ее точки остаются на месте.

Наметим идею доказательства этой теоремы при помощи так называемой сетки Мёбиуса.

Заметим, что 1) если при проективном преобразовании две точки остаются на месте, то прямая, через них проходящая, переходит в себя, и 2) если две прямые переходят в себя, то точка их пересечения остается на месте. Поэтому из того, что точки A, B, C, D плоскости Π^* остаются на месте, выходит последовательно, что и точки E, F, G, H, K, L и т. д.

тоже остаются на месте (рис. 81). Построение таких точек можно продолжить, соединяя уже полученные точки. Это — так называемая сетка Мёбиуса. Продолжая ее построение, можно сгущать ее до бесконечности. Можно показать, что совокупность ее узлов везде плотно покрывает всю плоскость. Поэтому, если еще предположить непрерывность проективного преобразования (что, впрочем, уже следует из его определения, но не очень просто), то окажется, что если при проективном преобразовании плоскости Π^* точки A, B, C, D остаются на месте, то и все точки плоскости Π^* остаются на месте.

Проективная геометрия. *Двумерной проективной геометрией* называется совокупность теорем о тех свойствах фигур проективной плоско-

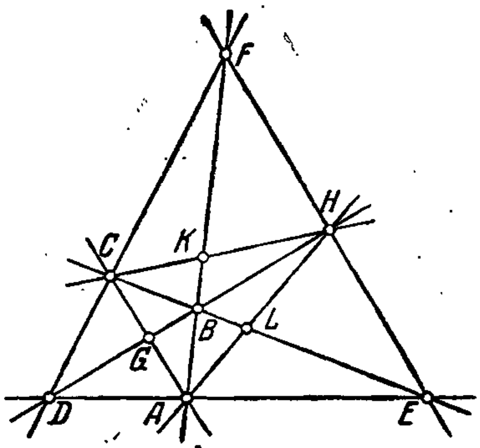


Рис. 81.

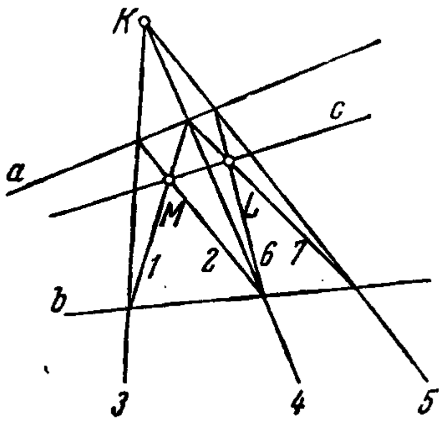


Рис. 82.

сти, т. е. обыкновенной плоскости, дополненной несобственными элементами, которые не изменяются при любых проективных преобразованиях.

Вот пример задачи проективной геометрии. Даны две прямые a и b и некоторая точка M (рис. 82). Построить прямую c, проходящую через точку M и точку пересечения прямых a и b, не используя самой этой точки пересечения (что может быть нужно, например, если эта точка очень далеко). Проведя через точку M две секущие *1* и *2* и затем прямые *3* и *4* через точки их пересечения с прямыми a и b, мы получаем точку K. Проведем через нее еще секущую *5* и прямые *6* и *7*, тогда, как можно доказать, прямая c, проходящая через точку L пересечения прямых *6* и *7* и точку M, и есть искомая.

Из теории конических сечений следует (рис. 83), что эллипс, гипербола и парабола суть перспективные проекции друг друга, и притом все они суть перспективные проекции окружности.

Если рассматривать перспективные проекции как проективные отображения проективных плоскостей P^* и $P^{*\prime}$ друг на друга, то, совместив эти плоскости, мы получим, что все эллипсы, гиперболы и параболы являются результатами проективных преобразований окружностей. Разница в том, что те проективные преобразования окружности, при которых в бесконечно удаленную прямую превращается прямая, не пересекающая эту окружность, суть эллипсы; если же в бесконечно удаленную прямую превращается прямая, касающаяся окружности, получается парабола, а если пересекающая — то гипербола (рис. 84).

Запись проективных преобразований формулами. Если на плоскости P^* взять обычные декартовы координаты, то, как можно показать, формулы

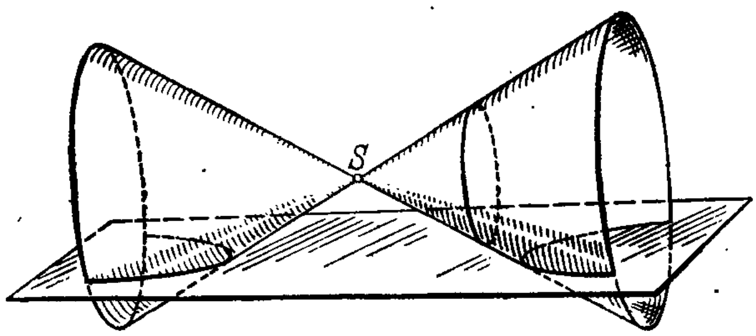


Рис. 83.

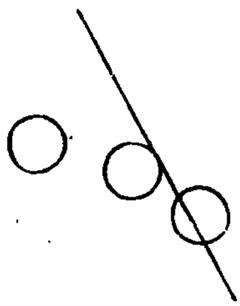


Рис. 84.

проективных преобразований плоскости имеют вид

$$x' = \frac{a_1x + b_1y + c_1}{a_3x + b_3y + c_3}, \quad y' = \frac{a_2x + b_2y + c_2}{a_3x + b_3y + c_3},$$

где определитель

$$\begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix} \neq 0,$$

и обратно.

Если для некоторой точки (x, y) знаменатели равны нулю, это значит, что ее образ (x', y') есть несобственная (бесконечно удаленная) точка. Уравнение

$$a_3x + b_3y + c_3 = 0$$

выражает ту прямую, которая при рассматриваемом проективном преобразовании переходит в несобственную (бесконечно удаленную прямую).

§ 14. ПРЕОБРАЗОВАНИЯ ЛОРЕНЦА

Вывод формул преобразования Лоренца для движений по прямой и в плоскости из положения о постоянстве скорости распространения света. В самом конце XIX в. в физике было обнаружено фундаментальное противоречие. Известный опыт Майкельсона, при котором измерялась скорость света (составляющая около 300 000 *км/сек*) в направлении движения Земли по ее орбите вокруг Солнца (скорость Земли около 30 *км/сек*) и в перпендикулярном к этому направлении, с неопровержимостью доказал, что все тела природы при своем движении, хотя бы в пустоте, как бы сокращаются в направлении движения. Теорию этого сокращения подробно исследовал голландский физик Лоренц. Оказалось, что это сокращение тем больше, чем скорость движущегося тела ближе к скорости света в пустоте, причем при скорости, равной скорости света, сокращение становится бесконечным. Лоренц выписал формулы этого сокращения. Однако вскоре физик Эйнштейн встал в этом вопросе на совсем иную точку зрения, к которой тогда был уже близок и Пуанкаре. Эйнштейн рассуждал так. Если считать, что для распространения света, как для обычного движения материального тела, верен закон сложения скоростей Галилея, то скорость света $c' = c + v$, где v — скорость наблюдателя, движущегося навстречу распространению света, а c — скорость света для неподвижного наблюдателя. Между тем из опыта Майкельсона следует, что $c' = c$. Закон $c' = c + v$ основывается на преобразовании

$$\begin{aligned} x' &= x + v_x t, \\ t' &= t, \end{aligned} \tag{25}$$

связывающем координату x некоторой точки по отношению к некоторой системе координат I с ее координатой x' по отношению к системе координат II, оси которой остаются параллельны осям системы I и которая движется в направлении оси Ox со скоростью v_x по отношению к системе I. Очевидно, эти-то формулы, как говорит Эйнштейн, и должны быть изменены.

Можно показать, как это, например, сделал недавно А. Д. Александров, что только из одного равенства скорости света в обеих координатных системах x, y, z, t и x', y', z', t' уже следует, что формулы преобразования от координат x, y, z, t к координатам x', y', z', t' — линейные и однородные, т. е. имеют вид

$$\begin{aligned} x' &= a_1 x + b_1 y + c_1 z + d_1 t, \\ y' &= a_2 x + b_2 y + c_2 z + d_2 t, \\ z' &= a_3 x + b_3 y + c_3 z + d_3 t, \\ t' &= a_4 x + b_4 y + c_4 z + d_4 t. \end{aligned} \tag{26}$$

Из других соображений можно показать, что определитель[1] их равен единице.

Если точка в системе I движется в произвольном заданном направлении прямолинейно и равномерно со скоростью света c, то $x = v_x t$, $y = v_y t$, $z = v_z t$ и $v_x^2 + v_y^2 + v_z^2 = c^2$, откуда

$$x^2 + y^2 + z^2 - c^2 t^2 = 0. \tag{27}$$

Но, согласно опыту Майкельсона, эта точка и в системе II должна двигаться с той же скоростью света c, и, следовательно, должно быть также

$$x'^2 + y'^2 + z'^2 - c^2 t'^2 = 0.$$

Формулы (26), таким образом, не произвольные линейные, однородные, с определителем, равным единице, а еще и такие, что если x, y, z, t удовлетворяют уравнению

$$x^2 + y^2 + z^2 - c^2 t^2 = 0,$$

то и результаты их преобразования x', y', z', t' также удовлетворяют

[1] См. главу XVI (том 3).

этому уравнению. Такие преобразования (26) называются *преобразованиями Лоренца*.

Рассмотрим сначала наиболее простой случай, когда точка движется вдоль оси Ox. В этом случае формулы (26) имеют вид

$$\begin{aligned} x' &= a_1x + d_1t, \\ t' &= a_2x + d_2t, \end{aligned} \tag{26'}$$

а уравнение (27)

$$x^2 - c^2t^2 = 0. \tag{27'}$$

Введем обозначение $ct = u$, тогда формулы (26′) и уравнение (27′) примут вид

$$\begin{aligned} x' &= a_1x + \frac{d_1}{c}u, \\ u' &= a_2cx + \frac{d_2c}{c}u \end{aligned} \tag{26$_1$}$$

и

$$x^2 - u^2 = 0. \tag{27$_1$}$$

Найдем в явном виде формулы (26_1). Рассмотрим x и u как декартовы прямоугольные координаты на плоскости, т. е. рассмотрим вопрос геометрически, причем будем считать, что формулы (26_1) суть формулы аффинного преобразования плоскости Oxu (определитель их, как было указано выше, равен единице). Преобразование это мы будем обозначать через L. Если по нашему предположению, из $x^2 - u^2 = 0$ следует $x'^2 - u'^2 = 0$, то это преобразование переводит крест прямых

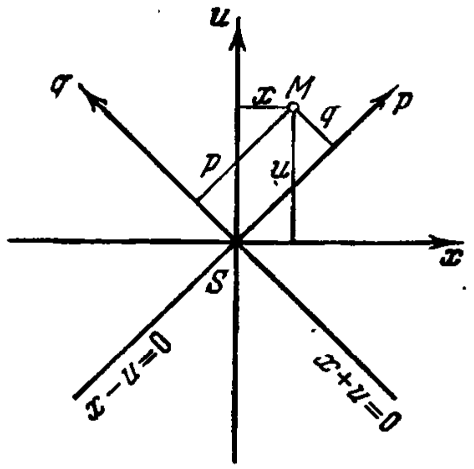


Рис. 85.

$$x^2 - u^2 = 0$$

в себя. Преобразование L, следовательно, — комбинация сжатия и растяжения с одинаковыми коэффициентами τ вдоль этих прямых.

Из рис. 85 получаем

$$\begin{aligned} x &= \frac{p}{\sqrt{2}} - \frac{q}{\sqrt{2}}, \\ u &= \frac{p}{\sqrt{2}} + \frac{q}{\sqrt{2}}. \end{aligned}$$

А так как после преобразования L, p и q перейдут в $p' = \frac{p}{\tau}$ и $q' = q\tau$, то

$$\begin{aligned} x'\sqrt{2} &= \frac{p}{\tau} - q\tau, \\ u'\sqrt{2} &= \frac{p}{\tau} + q\tau. \end{aligned}$$

Найдя p и q через x и u из первой пары уравнений, подставляя во вторую и упрощая, получим

$$x' = \frac{x - \frac{\tau^2 - 1}{\tau^2 + 1}ct}{\frac{2\tau}{\tau^2 + 1}}, \qquad t' = \frac{t - \frac{1}{c}\frac{\tau^2 - 1}{\tau^2 + 1}x}{\frac{2\tau}{\tau^2 + 1}},$$

или, полагая здесь $\frac{\tau^2 - 1}{\tau^2 + 1}c = v$, мы получаем

$$x' = \frac{x - vt}{\sqrt{1 - \left(\frac{v}{c}\right)^2}}, \qquad t = \frac{t - \frac{vx}{c^2}}{\sqrt{1 - \left(\frac{v}{c}\right)^2}},$$

— это и есть знаменитые *формулы Лоренца.*

Если взять здесь, в частности, $x = 0$, т. е. рассматривать движение начала координат системы I, то мы получим

$$x' = \frac{-vt}{\sqrt{1 - \left(\frac{v}{c}\right)^2}}, \quad t' = \frac{t}{\sqrt{1 - \left(\frac{v}{c}\right)^2}},$$

или $x' = -vt'$, откуда видно, что v есть скорость движения системы координат II по отношению к системе I.

Пусть, например, заданы две точки оси Ox с координатами x_1 и x_2 относительно системы I, расстояние между ними относительно системы I есть $r = |x_1 - x_2|$. Посмотрим, каково расстояние между ними для наблюдателя, связанного с системой II. Мы имеем

$$x_1' = \frac{x_1 - vt}{\sqrt{1 - \left(\frac{v}{c}\right)^2}}, \qquad x_2' = \frac{x_2 - vt}{\sqrt{1 - \left(\frac{v}{c}\right)^2}},$$

откуда

$$r' = |x_1' - x_2'| = \frac{|x_1 - x_2|}{\sqrt{1 - \left(\frac{v}{c}\right)^2}}.$$

Множитель $\sqrt{1 - \left(\frac{v}{c}\right)^2}$ есть, очевидно, как раз коэффициент сокращения Лоренца. Так как c очень велико, то при не очень больших v этот коэффициент весьма близок к единице и сокращения поэтому не заметно. Однако такие элементарные частицы, как электроны или позитроны, движутся часто со скоростями, сравнимыми со скоростью света, и поэтому при изучении их движений приходится учитывать эти обстоятельства, как говорят, приходится учитывать релятивистский эффект.

Перейдем теперь к следующему по сложности случаю, когда точка движется в плоскости Oxy. Для этого случая преобразования (26) будут иметь вид

$$\begin{aligned} x' &= a_1x + b_1y + d_1t, \\ y' &= a_2x + b_2y + d_2t, \\ t' &= a_3x + b_3y + d_3t, \end{aligned} \tag{26''}$$

где

$$\begin{vmatrix} a_1 & b_1 & d_1 \\ a_2 & b_2 & d_2 \\ a_3 & b_3 & d_3 \end{vmatrix} = 1,$$

а уравнение (27) будет

$$x^2 + y^2 - c^2t^2 = 0. \qquad (27'')$$

Это — формулы Лоренца для движения в плоскости Oxy.

Опять положим $ct = u$. Тогда преобразования (26″) напишутся так:

$$\begin{aligned} x' &= a_1x + b_1y + \frac{d_1}{c}u, \\ y' &= a_2x + b_2y + \frac{d_2}{c}u, \\ u' &= a_3cx + b_3cy + \frac{d_3c}{c}u, \end{aligned} \qquad (26_2)$$

причем определитель их будет снова равен единице, а уравнение (27″) примет более простой вид

$$x^2 + y^2 - u^2 = 0. \qquad (27_2)$$

Будем считать x, y, u декартовыми прямоугольными координатами точки обыкновенного трехмерного пространства и будем рассматривать формулы (26_2) как формулы аффинных преобразований этого пространства. Уравнение (27_2) выражает прямой круговой конус K с углом 90° при вершине (рис. 86).

С точки зрения этого геометрического (геометрического потому, что здесь мы $u = ct$ считаем просто пространственной координатой) толкования преобразования Лоренца, движения в плоскости суть не что иное, как все эквиаффинные (т. е. аффинные, не изменяющие объемов) преобразования этого пространства, которые преобразуют конус K в себя.

Рассмотрим некоторые такие специальные преобразования Лоренца:

1. Очевидно, что любой простой поворот пространства, как жесткого целого, вокруг оси конуса K на некоторый угол ω есть эквиаффинное преобразование пространства, преобразующее конус K в себя, т. е. некоторое специальное преобразование Лоренца. Будем его обозначать через ω.

2. Отражения пространства в любой плоскости π, проходящей через ось конуса K, суть, очевидно, также преобразования Лоренца. Мы будем их обозначать через π.

3. Наконец, рассмотрим еще следующее преобразование (рис. 87). Пусть v и w — какая-либо пара противоположных образующих конуса, P и Q — плоскости, касающиеся конуса по этим образующим. Плоскости эти взаимно перпендикулярны. Сделаем сжатие пространства к плоскости P и растяжение его с тем же коэффициентом от плоскости Q или наоборот. Например, сожмем пространство в три раза к плоскости P и растянем его тоже в три раза от плоскости Q. Такое преобразование пространства, очевидно, тоже аффинное и сохраняет величину всех объемов. Будем

его обозначать через L. Покажем, что это преобразование переводит конус K в себя. В силу того, что конус K имеет ось u своею осью вращения, можно всю фигуру повернуть так, чтобы образующие v и w лежали, например, в плоскости Sxu. Поэтому достаточно доказательство провести для этого случая.

Для доказательства пересечем конус K любой плоскостью R, параллельной плоскости Sxu. Уравнение такой плоскости $y = b$, где b — постоянная. Подставляя это значение в уравнение конуса K, получаем

$$x^2 - u^2 = -b^2.$$

Это уравнение гиперболы, для которой как раз прямые пересечения плоскости R с плоскостями P и Q суть асимптоты. Но так как для точек такой

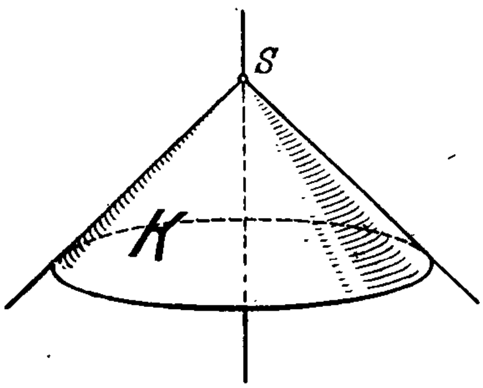


Рис. 86.

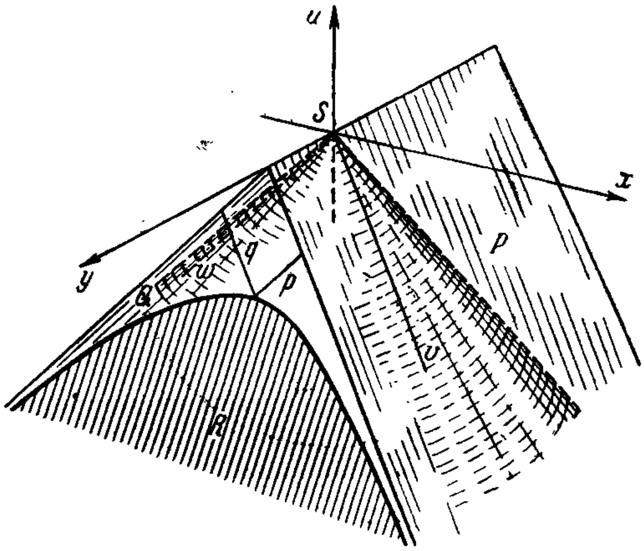


Рис. 87.

гиперболы характерно то, что произведение расстояний p и q до асимптот (т. е. до плоскостей P и Q) постоянно, при преобразовании L всякая точка такой гиперболы останется на этой же гиперболе и гипербола перейдет в себя. Но вся поверхность конуса K состоит из таких гипербол, и поэтому при преобразовании пространства L конус K преобразуется в себя. Это преобразование L поэтому — тоже преобразование Лоренца.

Так как при аффинных преобразованиях прямые переходят в прямые и пересекающиеся прямые переходят в пересекающиеся прямые, то связка S прямых при любом преобразовании Лоренца взаимно однозначно отображается на себя. Кроме того, при аффинных преобразованиях пространства всякая плоскость переходит в плоскость, поэтому при этих преобразованиях связки S на себя получается проективное преобразование этой связки. Если пересечь эту связку некоторой плоскостью Π, перпендикулярной к оси конуса K, которая, как целое, не участвует в рассмотренных преобразованиях Лоренца пространства, дополнить эту плоскость до проективной плоскости Π^* и следить за точками пересечения прямых связки S с плоскостью Π^*, то преобразования Лоренца, преобразуя связку, будут попутно давать некоторые проективные преобразования Λ

плоскости Π^*, переводящие круг α, по которому плоскость Π^* пересекается с внутренней частью конуса K, в себя. Для того чтобы разобраться в свойствах преобразований Лоренца, проще всего следить за вызываемыми ими проективными преобразованиями Λ круга α в себя.

Проективные преобразования круга в себя. Точку, исходящую из нее полупрямую и одну из полуплоскостей, отрезаемых этой прямой, мы будем называть «репером» плоскости Π^* (не путать с координатным репером, § 11). Покажем (рис. 88), что если взять два любых репера M и M', точки которых суть внутренние точки круга α, то можно при помощи преобразований L, ω и π преобразовать первый из этих реперов во второй.

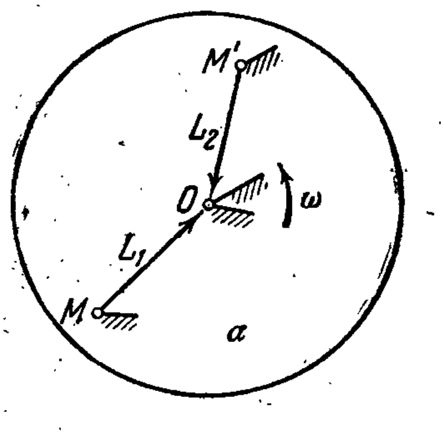


Рис. 88.

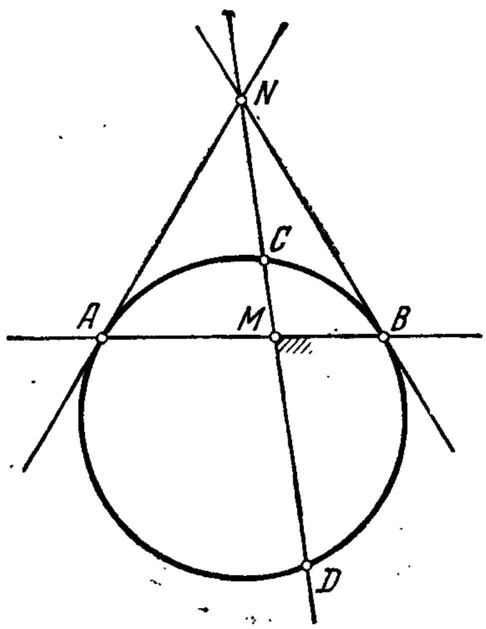


Рис. 89.

Для этого достаточно сделать преобразования $\Lambda = L_1 \cdot \omega \cdot L_2^{-1}$ (или преобразование $\Lambda = L_1 \cdot \omega \cdot \pi \cdot L_2^{-1}$). Преобразование L_1 приведет первый репер M в центр O круга α, преобразование ω повернет его как нужно и, наконец, преобразование L_2^{-1} совместит его со вторым репером M'.

Покажем, кроме того, что такое преобразование Λ, которое переводит какой-нибудь данный репер M в заданный репер M', только одно. Для этого заметим сначала, что если бы было два разных преобразования Λ_1 и Λ_2, переводящих репер M в репер M', то преобразование $\Lambda = \Lambda_1 \Lambda_2^{-1}$ было бы не тождественным преобразованием Λ и переводило бы репер M в себя. Достаточно поэтому показать, что если некоторое преобразование Λ переводит репер M в себя, то оно тождественное, т. е. оставляет все точки плоскости круга α на месте.

Докажем это. Пусть некоторое преобразование Λ переводит репер M в себя (рис. 89). Тогда оно переводит прямую AB этого репера в себя, но так как оно переводит в себя окружность круга α, то оно оставляет на месте точки A и B или меняет их местами. Последнее, однако, невозможно, так как оно переводит полупрямую репера в себя. Проведем в точках A и B касательные к кругу α. Они переходят в себя, так как если бы такая касательная сделалась секущей $A\bar{A}$, то обратное преобразование разные точки A и $\bar{A}$ окружности α преобразовало бы в одну точку A'. Но преобразования Λ проективные и, следовательно, взаимно однозначные. Раз при преобразовании Λ эти касательные переходят в себя, то точка N их пересечения остается на месте, и, следовательно, прямая MN переходит в себя. Из того, что полуплоскость репера M переходит в себя, заключаем, как выше, что точки C и D не меняются местами, а остаются на месте. Итак, при рассматриваемом проективном преобразовании Λ проективной плоскости Π^* четыре ее точки A, B, C, D, из которых никакие три не лежат на одной прямой, остаются на месте. По теореме единственности проективных преобразований это преобразование, следовательно, тождественное.

Дальше, в § 5 главы XVII (том 3), будет показано, что, используя выведенные сейчас свойства группы Лоренца, легко реализовать планиметрию Лобачевского, а если рассмотреть преобразования Лоренца для общего случая движения точки в пространстве, то и стереометрию Лобачевского, и тем самым показать непротиворечивость геометрии Лобачевского.

Мы видим, что теория преобразований Лоренца, проективная геометрия и теория перспективы и неэвклидова геометрия тесно связаны друг с другом. Оказывается, что с ними также в большой мере связана теория так называемых конформных преобразований в теории функций комплексного переменного, решающей такие важные задачи математической физики, как задачу распределения температур в нагретой пластинке, задачу об обтекании воздухом крыла самолета, задачу теории плоского электростатического поля, плоскую задачу теории упругости и многие другие.

ЗАКЛЮЧЕНИЕ

Аналитическая геометрия является совершенно необходимым математическим методом для изучения других отделов математики, механики, физики и других естественных наук. Поэтому ее изучают не только в университетах, но и во всех технических высших учебных заведениях, а также в некоторых техникумах. Ставится вопрос о включении довольно подробного изложения элементов аналитической геометрии в курс средней школы.

Различные координаты. Существенными частями идеи аналитической геометрии, как мы видели, является метод координат и рассмотрение уравнений, связывающих эти координаты. Кроме декартовых координат, рассматриваются различные другие координаты. Например, на плоскости можно, выбрав некоторую точку P (так называемый полюс) и некоторую исходящую из нее полупрямую (полярная ось), определять положение точки M длиной ρ полярного радиуса, идущего к ней из полюса, и величиной ω того угла, который этот радиус образует с полярной осью (рис. 90).

В частности, эллипс, гипербола или парабола, если за полюс взять фокус, а за полярную ось — полупрямую, идущую из него по оси симметрии в сторону, противоположную ближайшей вершине (рис. 91), имеют одно и то же уравнение

$$\rho=\frac{p}{1-\varepsilon\cos\omega},$$

где ε — эксцентриситет линии, а p — так называемый ее параметр. Это уравнение имеет большое значение в астрономии. Именно при помощи

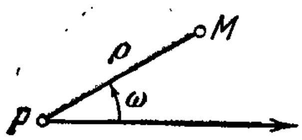


Рис. 90.

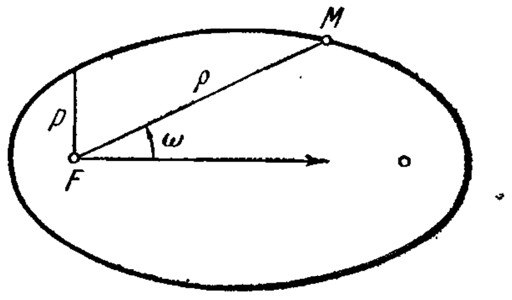


Рис. 91.

его исследования выводится из закона инерции и закона всемирного тяготения, что планеты обращаются вокруг Солнца по эллипсам.

Общеизвестны географические координаты — долгота и широта, которыми задается положение точки на сфере.

Аналогично можно ввести координатную сетку на любой поверхности, как это делают в дифференциальной геометрии (см. главу VII, том 2) и т. д.

Многомерная и бесконечномерная аналитическая геометрия. Алгебраическая геометрия. Казалось бы, что аналитическая геометрия к XIX в. прошла такой большой путь развития, описанный нами выше в самых общих чертах, и дала столько идей, что она должна была бы себя исчерпать, однако это не так. Как раз в настоящее время бурно развиваются две новые, весьма обширные ветви математики, продолжающие идеи аналитической геометрии,— так называемый *функциональный анализ* и общая *алгебраическая геометрия.* Правда, обе они только наполовину представляют собою прямое продолжение классической аналитической геометрии: в функциональном анализе много анализа, в алгебраической геометрии не мало теории функций и топологии.

Объясним, о чем идет речь. Еще в середине прошлого столетия начали рассматривать четырехмерную и вообще n-мерную аналитическую геометрию, т. е. стали изучать те вопросы алгебры, которые являются прямым обобщением алгебраических вопросов, изучаемых в двух- и трехмерной аналитической геометрии, но для случая, когда имеется 4 или n неизвестных. В самом конце XIX в. ряд выдающихся аналистов пришел к мысли, что для целей анализа и математической физики важно рассмотреть бесконечномерную аналитическую геометрию.

С первого взгляда может показаться, что если n-мерное и даже четырехмерное пространство кажутся какими-то надуманными математическими фикциями, то что же тогда говорить о бесконечномерном. Однако это не так. Рассуждения, касающиеся бесконечномерного пространства, вовсе не так трудны. Они составляют сейчас большую ветвь математики— функциональный анализ (см. главу XIX, том 3). В этом разделе математики одни из самых важных результатов даны в последнее время советскими учеными.

Любопытно, что бесконечномерная аналитическая геометрия имеет важнейшие практические приложения и играет фундаментальную роль в современной физике.

Что касается алгебраической геометрии, то это — более непосредственное продолжение обычной аналитической геометрии, которая сама в свою очередь есть лишь часть алгебраической геометрии. Алгебраическая геометрия может рассматриваться как та часть математики, которая занимается линиями, поверхностями и гиперповерхностями, выражаемыми в декартовых координатах алгебраическими уравнениями не только 1-й и 2-й степени, но и высших степеней. Оказалось, что в этих исследованиях выгодно рассматривать не только действительные, но и комплексные координаты, т. е. рассматривать всё в так называемом комплексном пространстве. Наиболее важные результаты в этой области были получены еще в прошлом столетии Риманом. Как на блестящий пример теоремы о линиях высших порядков, укажем на замечательный своею общностью результат И. Г. Петровского о числе овалов, на которые может разбиваться линия n-го порядка. И. Г. Петровский показал, что если p —число таких овалов, которые вовсе не лежат в других овалах либо лежат в четном числе овалов, а m — число тех овалов, которые лежат в нечетном числе овалов, и если рассматривать такие линии, в которых составляющие ее овалы ни сами себя, ни друг друга не пересекают (рис. 92), то

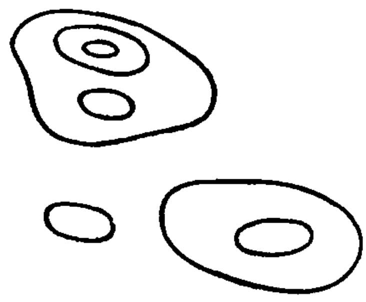


Рис. 92.

$$p - m \leqslant \frac{3n^2 - 6n}{8} + 1,$$

где n — порядок линии, т. е. степень того уравнения, которым она выражается.

Приведенный результат тем более важен, что до сих пор почти ничего не было известно об общем виде линий высших порядков. Это, пожалуй, одна из последних важных общих теорем, найденных в аналитической геометрии.

ЛИТЕРАТУРА

Популярные брошюры

Маркушевич А. И. Замечательные кривые. Гостехиздат, 1952.
Смогоржевский А. С. Метод координат. Гостехиздат, 1952.

Учебники для втузов

Ефимов Н. В. Краткий курс аналитической геометрии. Гостехиздат, 1954.
Привалов И. И. Аналитическая геометрия. Гостехиздат, 1954.

Учебники для университетов

Делоне Б. Н. и Райков Д. А. Аналитическая геометрия. Гостехиздат, т. I, 1948; т. II, 1949.
Мусхелишвили Н. И. Курс аналитической геометрии. Гостехиздат, 1947.

Глава IV.

АЛГЕБРА
(ТЕОРИЯ АЛГЕБРАИЧЕСКОГО УРАВНЕНИЯ)

§ 1. ВВЕДЕНИЕ

Что такое алгебра и каковы ее характерные черты, всем хорошо известно, потому что элементарные, но основные сведения об алгебре даются уже в средней школе. Алгебра характеризуется прежде всего своим методом, заключающимся в употреблении букв и буквенных выражений, над которыми производятся преобразования по определенным правилам. В элементарной алгебре под буквами подразумевают обыкновенные числа, поэтому правила преобразований буквенных выражений основываются на общих правилах действий над числами. [Например, сумма не зависит от порядка слагаемых, что записывается в алгебре так: $a + b = b + a$; при умножении суммы двух чисел можно помножить каждое из этих чисел в отдельности и затем сложить полученные произведения: $(a + b)c = ac + bc$ и т. д.].

Если проследить доказательство какой-нибудь алгебраической теоремы, то легко убедиться, что оно зависит только от тех правил, по которым производят действия над буквами, но вовсе не зависит от того, что эти буквы представляют.

Алгебраический метод, т. е. метод буквенных вычислений, пронизывает всю математику. Часто это находит свое выражение в том, что часть решения какого-нибудь математического вопроса представляет собою нечто иное, как более или менее сложную алгебраическую выкладку. Кроме того, в математике употребляют разные буквенные вычисления, в которых под буквами уже подразумевают не числа, а какие-либо другие объекты. При этом и правила действий над ними могут отличаться от правил элементарной алгебры. Например, в геометрии, механике и физике применяются векторы. Как известно, над векторами производят действия, правила которых отчасти такие же, как правила действий над числами, а отчасти существенно от них отличаются.

Значение алгебраического метода в современной математике и ее приложениях сильно возросло за последние десятилетия.

Во-первых, возросшие запросы техники заставляют доводить до численных результатов решение трудных задач математического анализа, что оказывается обычно возможным лишь после алгебраизации этих задач.

а это в свою очередь ставит новые, иногда трудные, задачи самой алгебре.

Во-вторых, некоторые вопросы анализа стали ясными и доступными только после того, как к ним начали применять алгебраические методы, основанные на глубоком обобщении (на случай, когда неизвестных бесконечно много) теории систем уравнений 1-й степени.

Наконец, высшие разделы алгебры нашли применение в современной физике, а именно: основные понятия квантовой механики, оказывается, выражаются при помощи сложных и неэлементарных алгебраических объектов.

Основные черты истории алгебры следующие.

Прежде всего надо отметить, что точка зрения на то, что такое алгебра и в чем состоит основная задача алгебры, два раза видоизменялась: один раз в первой половине прошлого века, а другой раз — в начале нашего столетия. Таким образом, под алгеброй в разное время понимали последовательно три довольно разные вещи. Этим история алгебры отличается от истории трех знаменитых исчислений: аналитической геометрии, дифференциального и интегрального, которые, выкованные руками их создателей — Ферма, Декарта, Ньютона, Лейбница и других, дальше бурно развивались и дополнялись иногда новыми большими разделами, однако в принципе сравнительно мало видоизменяли свое лицо.

В древние времена любое правило, найденное для решения некоторого класса математических задач, записывалось просто словами, так как буквенные обозначения еще не были придуманы. Самое слово «алгебра» производят от названия важнейшего сочинения хорезмского ученого IX в. Мухаммеда аль-Хорезми (см. главу I), в сочинениях которого приводятся первые общие правила для решения уравнений 1-й и 2-й степени. Однако введение самих буквенных обозначений обыкновенно связывают с именем Виета, который не только неизвестные, но и заданные величины начал обозначать буквами. Немало также сделал для развития буквенных обозначений Декарт, причем под буквами, конечно, понимались обыкновенные числа. С этого момента начинается собственно алгебра как наука о буквенных вычислениях, о преобразовании формул, составленных из букв, об алгебраических уравнениях и т. д., в отличие от арифметики, действующей всегда над конкретными числами. Только после этого даже и самые сложные математические соображения сделались легко обозримыми и доступными для изучения, так как, бросив взгляд на буквенную формулу, в большинстве случаев можно сразу видеть общее ее устройство — ее закон — и можно легко ее преобразовывать. В те времена всё в математике, что не есть ни геометрия, ни анализ бесконечно малых, называли алгеброй. Это — первая, так сказать виетовская, точка зрения на алгебру. Она особенно ясно выражена в известном «Введении в алгебру» члена Российской Академии наук, знаменитого Л. Эйлера,— книге, написанной в 60-х годах XVII в., т. е. 200 лет назад.

Алгебру Эйлер определяет как теорию вычислений с разными величинами. Его книга содержит в первой своей части теорию вычислений с целыми рациональными числами, обыкновенными дробями, квадратными и кубическими корнями, теорию логарифмов, прогрессии, теорию вычислений с многочленами, теорию биномиального ряда Ньютона и его приложений. Во второй части содержатся теория уравнений 1-й степени и их систем, теория квадратного уравнения и теория решения уравнений 3-й и 4-й степени в радикалах, а также обширный раздел, в котором разбираются способы решения различных неопределенных уравнений в целых числах. Например, доказывается невозможность решения уравнения Ферма $x^3 + y^3 = z^3$ в целых числах x, y, z.

В конце XVIII и начале XIX в. постепенно одна из задач алгебры, а именно теория решения алгебраических уравнений, в которой основной трудностью является решение алгебраического уравнения n-й степени с одной неизвестной

$$x^n + a_1x^{n-1} + a_2x^{n-2} + \ldots + a_{n-1}x + a_n = 0,$$

начала считаться центральной. Это произошло вследствие важности этой задачи для всей математики и ее приложений, а также ввиду трудности и глубины доказательств большинства связанных с нею теорем.

Всякому известна общая формула

$$x = -\frac{p}{2} \pm \sqrt{\frac{p^2}{4} - q},$$

при помощи которой решается любое квадратное уравнение. Итальянские алгебраисты XVI в. нашли аналогичные, хотя и более сложные общие правила для решения любых уравнений 3-й и 4-й степени. Дальнейшие исследования в этом деле для уравнений более высоких степеней, однако, натолкнулись на непреодолимые трудности. Величайшие математики XVI, XVII, XVIII и начала XIX в. (Тарталья, Кардан, Декарт, Ньютон, Эйлер, Даламбер, Чирнгаузен, Безу, Лагранж, Гаусс, Абель, Галуа, Лобачевский, Штурм и другие) создали грандиозный комплекс теорем и методов, связанных с этим вопросом. В двухтомной алгебре Серре (создавшей в свое время эпоху, так как в ней впервые излагалась вершина теории алгебраических уравнений — теория Галуа), появившейся в середине XIX в., ровно через сто лет после алгебры Эйлера, алгебра уже определяется как теория алгебраических уравнений. Это — вторая точка зрения на то, что такое алгебра.

Во второй половине прошлого столетия, исходя из идей Галуа, связанных с теорией алгебраического уравнения, были глубоко развиты теория групп[1] и теория алгебраических чисел (в создании которой большую роль сыграл русский математик Е. И. Золотарев).

В этот же второй период в связи с той же задачей решения алгебраического уравнения, а также с теорией алгебраических многообразий

[1] См. главу XX (том 3).

высших степеней, изучавшихся тогда в аналитической геометрии, развился в различных направлениях алгебраический аппарат — теория определителей и матриц, алгебраическая теория квадратичных форм и линейных преобразований и особенно теория инвариантов. В течение почти всей второй половины XIX в. теория инвариантов была одной из центральных тем алгебраических исследований. В свою очередь развитие теории групп и теории инвариантов оказало в этот период большое влияние на развитие геометрии[1].

Новая, третья точка зрения на то, что такое алгебра, появилась в основном в связи со следующим. Во второй половине прошлого века в механике, физике и самой математике начали все чаще и чаще изучаться такие объекты, для которых естественно рассматривать действия сложения и вычитания, а иногда и умножения и деления, однако действия эти подчинены уже другим законам, чем для рациональных чисел.

Мы уже говорили о векторах. Другие виды такого рода величин с другими законами операций над ними можно здесь только назвать: это матрицы, тензоры, спиноры, гиперкомплексные числа и т. д. Все эти величины обозначают буквами, но для разных видов этих величин законы операций отличаются друг от друга. Если для некоторого множества объектов (обозначаемых буквами) заданы некоторые операции и законы (правила), которым удовлетворяют эти операции, то говорят, что задана некоторая алгебраическая система. Третья точка зрения на то, что такое алгебра, состоит в том, что целью алгебры является изучение разных алгебраических систем. Это так называемая аксиоматическая или абстрактная алгебра. Абстрактная она потому, что на данном этапе все равно, что именно обозначают буквы в рассматриваемой алгебраической системе, нам важно только, каким аксиомам (законам) удовлетворяют рассматриваемые в системе операции. Аксиоматической же ее называют потому, что она строится, исключительно исходя из тех аксиом, которые положены в ее основание. Вернулись снова, но уже на высшей ступени, как бы к первой, виетовской, точке зрения на алгебру, что алгебра есть теория буквенных вычислений. Что же понимать под буквами — все равно, важны только те законы, которым удовлетворяют производимые над буквами операции, причем интересны, конечно, только такие алгебраические системы, которые имеют большое значение или в самой математике, или в ее приложениях.

Большой алгебраический материал, собранный в предыдущий период, послужил фактической основой для построения современной абстрактной алгебры.

В 30-х годах XX в. появился известный курс Ван дер Вардена «Современная алгебра», который сыграл большую роль в деле пропаганды этой третьей точки зрения на то, что такое алгебра. В таком же направлении написан курс алгебры А. Г. Куроша.

---

[1] См. главу XVII (том 3).

В настоящем столетии алгебра получила глубочайшие приложения к геометрии (к топологии и к теории групп Ли) и, как мы об этом уже говорили выше, к современной физике (функциональный анализ и квантовая механика).

В самое последнее время стали особенно важными вопросы о механизации алгебраических вычислений при помощи различных математических вычислительных машин, особенно при помощи быстродействующих электронных машин. Вопросы, связанные с этой машинной математикой, поставили новые своеобразные задачи алгебре.

В настоящей книге, кроме этой главы, алгебре посвящены еще две: линейная алгебра (глава XVI, том 3) и теория групп и других алгебраических систем (глава XX, том 3).

§ 2. АЛГЕБРАИЧЕСКОЕ РЕШЕНИЕ УРАВНЕНИЯ

Алгебраическим уравнением n-й степени с одной неизвестной называется уравнение вида

$$x^n + a_1x^{n-1} + a_2x^{n-2} + \ldots + a_{n-1}x + a_n = 0,$$

где $a_1, a_2, \ldots, a_n$ — некоторые заданные коэффициенты[1].

Уравнения 1-й и 2-й степени. Если уравнение 1-й степени, то оно имеет вид

$$x + a = 0$$

и решается сразу

$$x = -a.$$

Уравнение 2-й степени

$$x^2 + px + q = 0$$

было решено в глубокой древности. Оно решается просто: переносят q с обратным знаком в правую часть и, прибавляя затем к обеим частям $\frac{p^2}{4}$, получают

$$x^2 + px + \frac{p^2}{4} = \frac{p^2}{4} - q.$$

Но

$$x^2 + px + \frac{p^2}{4} = \left(x + \frac{p}{2}\right)^2,$$

следовательно

$$x + \frac{p}{2} = \pm\sqrt{\frac{p^2}{4} - q},$$

откуда и получается известная формула решения квадратного уравнения

$$x = -\frac{p}{2} \pm \sqrt{\frac{p^2}{4} - q}.$$

[1] Мы предполагаем, что все члены уравнения перенесены в левую часть и что оно поделено на коэффициент при высшей степени неизвестной.

Уравнение 3-й степени. Совершенно иначе было с уравнениями выше 2-й степени. Уже общее уравнение 3-й степени потребовало совершенно не очевидных соображений и не поддавалось всем усилиям математиков древности. Оно было решено лишь в начале 1500-х годов, в эпоху Возрождения в Италии, итальянским математиком Сципио дель Ферро. Ферро своего открытия, по обычаю того времени, не опубликовал, но сообщил одному из своих учеников. Уже после смерти Ферро ученик этот вызвал на соревнование одного из крупнейших итальянских математиков — Тарталья — и предложил ему для решения ряд уравнений 3-й степени. Тарталья (1500—1557) вызов принял и за 8 дней до конца состязания нашел способ решить любое кубическое уравнение вида $x^3 + px + q = 0$.

За 2 часа он решил все задачи противника. Профессор математики и физики в Милане Кардан (1501—1576), узнав об открытии Тарталья, начал умолять Тарталья сообщить ему свою тайну. Тарталья, в конце концов, согласился, но с условием, чтобы Кардан хранил его способ в глубоком секрете. Кардан нарушил обещание и опубликовал результат Тарталья в своем сочинении «Великое искусство» («Ars magna»).

Формула для решения кубического уравнения с тех пор называется формулой Кардана, хотя ее по справедливости надо было бы называть формулой Тарталья.

Вывод формулы Кардана получается так.

Во-первых, решение общего кубического уравнения

$$y^3 + ay^2 + by + c = 0 \tag{1}$$

легко свести к решению кубического уравнения вида

$$x^3 + px + q = 0, \tag{2}$$

не содержащего члена с квадратом неизвестной. Для этого достаточно положить $y = x - \frac{a}{3}$. Действительно, подставив это выражение в уравнение (1) и раскрывая скобки, получаем

$$\left(x - \frac{a}{3}\right)^3 + a\left(x - \frac{a}{3}\right)^2 + b\left(x - \frac{a}{3}\right) + c = x^3 - 3x^2\frac{a}{3} + \cdots + ax^2 + \cdots,$$

где точками обозначены члены, в которые x входит в 1-й степени или вовсе не входит. Мы видим, что члены, содержащие x^2, взаимно уничтожаются.

Пусть теперь задано уравнение

$$x^3 + px + q = 0.$$

Положим $x = u + v$, т. е. вместо одной неизвестной введем две u и v и тем самым превратим всю задачу в задачу с двумя неизвестными. Мы будем иметь

$$(u + v)^3 + p(u + v) + q = 0,$$

или

$$u^3 + v^3 + q + (3uv + p)(u + v) = 0.$$

Какова бы ни была сумма двух чисел $u+v$, всегда можно потребовать, чтобы произведение их uv равнялось любой наперед заданной величине. Если $u+v=A$, а мы хотим, чтобы $uv=B$, то, так как $v=A-u$, мы получаем

$$u(A-u)=B,$$

так что достаточно, чтобы u было корнем квадратного уравнения

$$u^2-Au+B=0,$$

а всякое квадратное уравнение имеет действительный или комплексный корень, получаемый по известной формуле. В нашем случае $u+v$ равняется искомому корню x нашего кубического уравнения; потребуем, чтобы

$$uv=-\frac{p}{3},$$

т. е. чтобы было $3uv+p=0$. При таком выборе u и v мы получим

$$\begin{aligned} u^3+v^3+q&=0,\\ 3uv+p&=0. \end{aligned} \tag{3}$$

При этом, если мы найдем числа u и v, удовлетворяющие этой системе уравнений, то число $x=u+v$ будет корнем нашего уравнения.

Из системы (3) легко образовать некоторое квадратное уравнение, корнями которого будут u^3 и v^3. Действительно, она дает

$$\begin{aligned} u^3+v^3&=-q,\\ u^3v^3&=-\frac{p^3}{27}, \end{aligned}$$

и, следовательно, по использованной уже выше теореме u^3 и v^3 суть корни квадратного уравнения

$$z^2+qz-\frac{p^3}{27}=0.$$

Решая его по обычной формуле, мы получаем

$$u^3=-\frac{q}{2}+\sqrt{\frac{q^2}{4}+\frac{p^3}{27}},\qquad v^3=-\frac{q}{2}-\sqrt{\frac{q^2}{4}+\frac{p^3}{27}},$$

и, следовательно,

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

— это и есть формула Кардана.

Уравнение 4-й степени. Вскоре после решения кубического уравнения было решено Феррари (1522—1565) и общее уравнение 4-й степени. При этом, если для решения уравнения 3-й степени понадобилось предварительное решение вспомогательного квадратного уравнения

$$z^2+qz-\frac{p^3}{27}=0,$$

где $z=u^3$ или v^3, то решение уравнения 4-й степени аналогично опирается на предварительное решение некоторого вспомогательного кубического уравнения.

Способ Феррари состоит в следующем. Пусть дано общее уравнение 4-й степени

$$x^4 + ax^3 + bx^2 + cx + d = 0.$$

Напишем его так:

$$x^4 + ax^3 = -bx^2 - cx - d$$

и прибавим к обеим частям $\frac{a^2x^2}{4}$, тогда слева получим полный квадрат

$$\left(x^2 + \frac{ax}{2}\right)^2 = \left(\frac{a^2}{4} - b\right)x^2 - cx - d.$$

Прибавим теперь к обеим частям уравнения еще члены

$$\left(x^2 + \frac{ax}{2}\right)y + \frac{y^2}{4},$$

где y — новая переменная, на которую мы далее наложим нужное нам условие, тогда слева получим опять полный квадрат

$$\left(x^2 + \frac{ax}{2} + \frac{y}{2}\right)^2 = \left(\frac{a^2}{4} - b + y\right)x^2 + \left(\frac{ay}{2} - c\right)x + \left(\frac{y^2}{4} - d\right). \qquad (4)$$

Таким образом, мы свели вопрос к задаче с двумя неизвестными.

Справа в равенстве (4) имеем квадратный трехчлен от x, коэффициенты которого зависят от y. Подберем y так, чтобы этот трехчлен был квадратом двучлена $\alpha x + \beta$ первой степени.

Для того чтобы квадратный трехчлен $Ax^2 + Bx + C$ был квадратом двучлена $\alpha x + \beta$, достаточно, чтобы

$$B^2 - 4AC = 0.$$

Действительно, если $B^2 - 4AC = 0$, то

$$Ax^2 + Bx + C = (\sqrt{A}x + \sqrt{C})^2,$$

т. е.

$$Ax^2 + Bx + C = (\alpha x + \beta)^2,$$

где

$$\alpha = \sqrt{A}, \quad \beta = \sqrt{C}.$$

Следовательно, если подобрать y так, чтобы

$$\left(\frac{ay}{2} - c\right)^2 - 4\left(\frac{a^2}{4} - b + y\right)\left(\frac{y^2}{4} - d\right) = 0,$$

то правая часть уравнения (4) будет полным квадратом $(\alpha x + \beta)^2$. Раскрывая скобки, получаем кубическое уравнение относительно y

$$y^3 - by^2 + (ac - 4d)y - [d(a^2 - 4b) + c^2] = 0.$$

Решив это вспомогательное кубическое уравнение (например, по формуле Кардана), мы через его корень y_0 найдем α и β и будем иметь

$$\left(x^2 + \frac{ax}{2} + \frac{y_0}{2}\right)^2 = (\alpha x + \beta)^2,$$

откуда

$$x^2 + \frac{ax}{2} + \frac{y_0}{2} = \alpha x + \beta \text{ или } x^2 + \frac{ax}{2} + \frac{y_0}{2} = -\alpha x - \beta.$$

Из этих двух квадратных уравнений мы и найдем все четыре корня заданного уравнения 4-й степени.

Так были решены итальянскими математиками в 1500-х годах алгебраические уравнения 3-й и 4-й степени.

Успех итальянских математиков произвел очень большое впечатление. То был первый случай, когда наука нового времени превзошла достижения древних. До того, в течение всех средних веков, ставили себе целью хотя бы понять сочинения древних, а тут, наконец, решили вопросы, которых древним решить не удалось. И это было в 1500-х годах, т. е. за столетие до изобретения новых исчислений: аналитической геометрии, дифференциального и интегрального, которые окончательно утвердили превосходство новой науки над старой. Не было после этого крупного математика, который не пытался бы продолжить достижения итальянцев и аналогично решить в радикалах уравнения 5-й, 6-й и высших степеней.

Выдающемуся алгебраисту XVII в. Чирнгаузену (1651—1708) даже показалось, что он, наконец, нашел общий метод решения. Метод его был основан на преобразовании уравнения к более простому, но это преобразование само требовало решения некоторых вспомогательных уравнений. Впоследствии, при более глубоком рассмотрении, оказалось, что метод преобразований Чирнгаузена действительно дает решение уравнений 2-й, 3-й и 4-й степени, но уже для уравнения 5-й степени требует предварительного решения вспомогательного уравнения 6-й степени, которое в свою очередь неизвестно как решить.

Разложение многочлена на множители и формулы Виета. Если принять без доказательства так называемую основную теорему алгебры[1], что всякое уравнение

$$f(x) = 0,$$

где

$$f(x) = x^n + a_1x^{n-1} + \ldots + a_n$$

многочлен от x любой данной n-й степени, а коэффициенты $a_1, a_2, \ldots, a_n$ — заданные действительные или комплексные числа, имеет хотя бы один действительный или комплексный корень, и принять во внимание, что все вычисления с комплексными числами производятся по тем же законам, как с рациональными числами, то легко показать, что многочлен $f(x)$ можно представить, и притом только одним способом, в виде произведения множителей 1-й степени

$$f(x) = (x - a)(x - b)\ldots(x - l),$$

где $a, b, \ldots, l$ — некоторые действительные или комплексные числа.

В самом деле, пусть a — корень $f(x)$; будем делить $f(x)$ на $x - a$;

[1] Доказательство основной теоремы алгебры трудно и оно было дано гораздо позже. У нас ей посвящен § 3. Но правильность ее предполагали уже задолго до строгого ее доказательства.

так как делитель 1-й степени, то остатком будет постоянное число R, т. е. мы будем иметь тождество

$$f(x) = (x-a) f_1(x) + R,$$

где $f_1(x)$—многочлен $(n-1)$-й степени, а R — постоянное. Подставляя сюда вместо x число a, мы получим

$$f(a) = (a-a) f_1(a) + R = R.$$

Но так как a — корень $f(x)$, то $f(a) = 0$ и, следовательно, $R = 0$, т. е. многочлен всегда делится нацело на $x - a$, где a — корень этого многочлена. Итак

$$f(x) = (x-a) f_1(x).$$

Но если верна основная теорема алгебры, то в свою очередь и многочлен $f_1(x)$ имеет некоторый корень b, и мы аналогично получим

$$f_1(x) = (x-b) f_2(x),$$

где многочлен $f_2(x)$ уже $(n-2)$-й степени и т. д. Разложение это, как легко показать, единственное.

Всякий многочлен n-й степени $f(x)$ имеет в этом смысле n и только n корней $a, b, c, \ldots, l$. Причем все эти корни могут быть различными, но могут быть среди них и одинаковые. Тогда мы говорим, что соответственный корень многочлена $f(x)$ кратный, с такою-то кратностью.

Перемножая скобки

$$(x-a)(x-b)(x-c)\ldots(x-l)$$

и сравнивая коэффициенты при одинаковых степенях x, непосредственно убеждаемся, что

$$\begin{aligned} -a_1 &= a + b + c + \ldots + l, \\ a_2 &= ab + ac + \ldots + kl, \\ -a_3 &= abc + abd + \ldots, \\ &\ldots\ldots\ldots\ldots\ldots \\ \pm a_n &= abc \ldots l \end{aligned}$$

— это формулы Виета.

Теорема о симметрических многочленах. Формулы Виета суть многочлены от n букв $a, b, \ldots, l$, которые не изменяются при любых перестановках этих букв. Действительно, $a + b + \ldots + k + l = b + a \ldots + k + l$ и т. д. Вообще любые такие многочлены от n букв, которые не изменяются при любых перестановках этих букв, называются симметрическими многочленами от n букв. Например, $5x^2 + 5y^2 - 7xy$ — симметрический многочлен от x и y. Можно доказать теорему, что всякий целый симметрический многочлен от n букв с любыми коэффициентами $A, B, \ldots$ может быть выражен цело-рационально (т. е. при помощи действий сложения, вычитания и умножения) через коэффициенты $A, B, \ldots$ и многочлены Виета от рассматриваемых букв. В случае если $a, b, \ldots, l$—корни уравнения n-й степени $x^n + a_1x^{n-1} + \ldots + a_n = 0$, то всякий симметрический многочлен от $a, b, \ldots, l$ с любыми коэффициентами $A, B, \ldots$ может быть, следовательно, выражен цело-рационально через эти коэффициенты $A, B, \ldots$ и коэффициенты $a_1, a_2, \ldots, a_n$ уравнения. Это — так называемая основная теорема о симметрических многочленах.

Работы Лагранжа. Знаменитый французский математик Лагранж в большой работе (имеющей более 200 страниц), появившейся в 1770—1771 гг., «Размышления о решении алгебраических уравнений» рассмотрел критически все известные до него способы решения уравнений 2-й, 3-й и 4-й степени и показал, что успех их решения был основан всегда на обстоятельствах, которые для уравнений 5-й и высших степеней уже не имеют места. Со времени Ферро до этой работы Лагранжа прошло более двух с половиной столетий, и никто за этот длинный промежуток времени не сомневался в возможности решить уравнения 5-й и высших степеней в радикалах, т. е. в возможности найти формулы, содержащие только действия сложения, вычитания, умножения, деления и извлечения корней с целыми положительными показателями, которые выражали бы корни уравнения через его коэффициенты,— формулы, подобные тем, при помощи которых в древности было решено квадратное уравнение и в 1500-х годах были решены итальянцами уравнения 3-й и 4-й степени. Считалось только, что никак не удается найти верный и, повидимому, глубоко скрытый путь к этому решению.

Лагранж в своем мемуаре (стр. 305, т. 3 Полного собрания сочинений) говорит: «Задача решения (в радикалах) уравнений, степени которых выше четырех, одна из тех, которую не удается решить, хотя ничто и не доказывает невозможности такого решения», а на стр. 307 он добавляет: «Из наших рассуждений следует, что весьма сомнительно, чтобы методы, которые мы рассмотрели, могли дать полное решение уравнений пятой степени».

В своих исследованиях Лагранж ввёл в рассмотрение выражения

$$a + \varepsilon b + \varepsilon^2 c + \ldots + \varepsilon^{n-1} l$$

от корней $a, b, \ldots, l$ уравнения, где ε — какой-нибудь корень n-й степени из единицы[1], установив, что именно такие выражения тесно связаны с решением уравнений в радикалах. Выражения эти сейчас называются «резольвентами Лагранжа».

Кроме того, Лагранж заметил, что большое значение в теории решения уравнения в радикалах имеет теория перестановок корней уравнения. Он даже высказал мысль, что теория перестановок является «истинной

[1] Т. е. такое комплексное число, которое, будучи возведено в n-ю степень, равно единице. Например, корень кубический из единицы может иметь значения 1, $-\frac{1}{2} + \frac{\sqrt{3}}{2} i$, $-\frac{1}{2} - \frac{\sqrt{3}}{2} i$, где $i = \sqrt{-1}$ (см. стр. 267). Действительно, $\left(-\frac{1}{2} + \frac{\sqrt{3}}{2} i\right)^3 = -\frac{1}{8} - \frac{3}{8}\sqrt{3}\, i + \frac{9}{8} + \frac{3\sqrt{3}}{8} i = 1$, и аналогично $\left(-\frac{1}{2} - \frac{\sqrt{3}}{2} i\right)^3 = 1$.

философией всего вопроса», в чем он был вполне прав, как это показали позднейшие исследования Галуа.

Теперь у Лагранжа выводы решений уравнений 2-й, 3-й и 4-й степени получились не такие, как у итальянцев,— в каждом случае по-своему и из каких-то сложных и как бы случайно найденных преобразований, а вполне стройные и выведенные из одной общей идеи при помощи единообразного метода теории симметрических многочленов, теории подстановок и теории резольвент.

Рассмотрим, для примера, решение способом Лагранжа общего уравнения 4-й степени

$$x^4 + mx^3 + nx^2 + px + q = 0.$$

Пусть корни этого уравнения a, b, c, d. Рассмотрим резольвенту

$$a + b - c - d,$$

т. е.

$$a + \varepsilon c + \varepsilon^2 b + \varepsilon^3 d,$$

где $\varepsilon = -1$. Переставляя в ней a, b, c, d всеми $1\cdot 2\cdot 3\cdot 4 = 24$ различными способами, мы получаем всего шесть различных выражений:

$$\begin{aligned} &a + b - c - d,\\ &a + c - b - d,\\ &a + d - c - b,\\ &c + d - a - b,\\ &b + d - a - c,\\ &b + c - a - d. \end{aligned} \tag{5}$$

Уравнение 6-й степени, корнями которого являются эти шесть выражений, будет, следовательно, иметь коэффициенты, которые не изменяются от всех 24 перестановок a, b, c, d, так как любая из 24 перестановок может только переставить эти выражения друг с другом, а коэффициенты рассматриваемого уравнения 6-й степени не зависят от порядка, в котором мы берем его корни. Таким образом, эти коэффициенты — симметрические многочлены от a, b, c, d. Но тогда, в силу основной теоремы о симметрических многочленах, коэффициенты эти выражаются цело-рационально через коэффициенты m, n, p, q уравнения. Кроме того, так как выражения (5) — попарно обратных знаков, то это уравнение 6-й степени будет содержать только члены четных степеней. Действительно, если выражения (5) обозначить через α, β, γ, $-\alpha$, $-\beta$, $-\gamma$, то левая часть рассматриваемого уравнения 6-й степени будет равна

$$(y-\alpha)(y+\alpha)(y-\beta)(y+\beta)(y-\gamma)(y+\gamma) = (y^2-\alpha^2)(y^2-\beta^2)(y^2-\gamma^2).$$

Непосредственное вычисление дает уравнение 6-й степени

$$y^6 - (3m^2 - 8n)y^4 + 3(m^4 - 16m^2 n - 16n^2 + 16mp - 64q)y^2 - \\ - (m^2 - 4m + 8p)^2 = 0.$$

Полагая $y^2 = t$, получаем кубическое уравнение для t, и если t', t'', t''' — его корни, то

$$a + b - c - d = \sqrt{t'},$$
$$a + c - b - d = \sqrt{t''},$$
$$a + d - b - c = \sqrt{t'''}.$$

Кроме того, мы еще имеем

$$a + b + c + d = -m.$$

Складывая эти уравнения после предварительного умножения их на 1, 1, 1, 1 или 1, —1, —1, 1, или —1, 1, —1, 1, или —1, —1, 1, 1, мы получаем

$$a = \frac{1}{4}(-m + \sqrt{t'} + \sqrt{t''} + \sqrt{t'''}),$$
$$b = \frac{1}{4}(-m + \sqrt{t'} - \sqrt{t''} - \sqrt{t'''}),$$
$$c = \frac{1}{4}(-m - \sqrt{t'} + \sqrt{t''} - \sqrt{t'''}),$$
$$d = \frac{1}{4}(-m - \sqrt{t'} - \sqrt{t''} + \sqrt{t'''}).$$

Таким образом, решение уравнения 4-й степени сведено к решению кубического уравнения. Аналогично решаются и уравнения 3-й и 2-й степени.

Лагранж в теории алгебраического уравнения достиг многого. Однако даже и после его упорных усилий вопрос о решении в радикалах алгебраических уравнений, степень которых выше четвертой, остался открытым. Этот вопрос, над которым бесплодно работали почти три столетия, как выразился Лагранж, «был как бы вызовом человеческому уму».

Открытие Абеля. Каково же было удивление всех математиков, когда в 1824 г. вышла в свет работа молодого гениального норвежца Абеля (1802—1829), в которой дано доказательство того, что если коэффициенты уравнения $a_1, a_2, \ldots, a_n$ считать просто буквами, то не существует *никакого* радикального выражения, составленного из этих коэффициентов, которое было бы корнем соответственного уравнения, если степень его $n \geqslant 5$. Итак, за три столетия усилия величайших математиков всех стран решить в радикалах уравнение 5-й или высшей степени потому не увенчались успехом, что эта задача просто не имеет решения.

Известна такая формула для уравнений 2-й степени; имеются, как мы видели, аналогичные формулы для уравнений 3-й и для уравнений 4-й степени, а уже для уравнения 5-й или более высокой степени никакой такой формулы нет.

Доказательство Абеля трудное, и мы его здесь приводить не будем.

Теория Галуа. Однако и это было еще не все. Самое замечательное в теории алгебраического уравнения еще оставалось впереди. Дело в том, что есть сколько угодно частных видов уравнений всех степеней, которые решаются в радикалах, и как раз уравнений, важных во многих приложениях. Таковыми являются, например, двучленные уравнения $x^n = A$.

Абель нашел другой очень широкий класс таких уравнений, так называемые циклические уравнения и еще более общие «абелевы» уравнения. Гаусс по поводу задачи построения циркулем и линейкой правильных многоугольников подробно рассмотрел так называемое уравнение деления круга, т. е. уравнение вида

$$x^{p-1} + x^{p-2} + \ldots + x + 1 = 0,$$

где p — простое число, и показал, что оно всегда может быть сведено к решению цепи уравнений низших степеней, причем нашел условия, необходимые и достаточные для того, чтобы такое уравнение решалось в квадратных радикалах. (Необходимость этих условий была строго обоснована только Галуа.)

Итак, после работ Абеля положение было следующее: хотя, как это показал Абель, общее уравнение, степень которого выше четвертой, вообще говоря, не решается в радикалах, однако есть сколько угодно различных частных уравнений любых степеней, которые все же решаются в радикалах. Весь вопрос о решении уравнений в радикалах был поставлен этими открытиями на совсем новую почву. Стало ясно, что надо искать, каковы все те уравнения, которые решаются в радикалах, или, иначе говоря, каково условие, необходимое и достаточное для того, чтобы уравнение решалось в радикалах. Этот вопрос, ответ на который давал в некотором смысле окончательное выяснение всей задачи, решил гениальный французский математик Эварист Галуа.

Галуа (1811—1832) погиб в возрасте 20 лет на дуэли и в последние два года своей жизни не мог посвящать много времени занятиям математикой, так как был увлечен бурным вихрем политической жизни времен революции 1830 г., сидел в тюрьме за свои выступления против реакционного режима Людовика-Филиппа и т. п. Тем не менее за свою короткую жизнь Галуа сделал в разных частях математики открытия, далеко опередившие его время, и, в частности, дал самые замечательные из имеющихся результатов в теории алгебраических уравнений. В небольшой работе «Мемуар об условиях разрешимости уравнений в радикалах», оставшейся в его рукописях после его смерти и впервые обнародованной Лиувиллем лишь в 1846 г., Галуа, исходя из самых простых, но глубоких соображений, наконец, распутал весь клубок трудностей, сосредоточенных вокруг теории решения уравнений в радикалах,— трудностей, над которыми безуспешно бились до того величайшие математики. Успех Галуа был основан на том, что он первый применил в теории уравнений ряд чрезвычайно важных новых общих понятий, впоследствии сыгравших большую роль во всей математике в целом.

Рассмотрим теорию Галуа для частного случая, а именно того, когда коэффициенты $a_1, a_2, \ldots, a_n$ заданного уравнения n-й степени

$$x^n + a_1x^{n-1} + \ldots + a_{n-1}x + a_n = 0 \qquad (6)$$

— рациональные числа. Случай этот особенно интересен и содержит в себе по существу уже все трудности общей теории Галуа. Мы будем, кроме того, предполагать, что все корни $a, b, c,\ldots$ рассматриваемого уравнения различны.

Галуа начинает с того, что, подобно Лагранжу, рассматривает некоторое выражение 1-й степени относительно $a, b, c,\ldots$

$$V = Aa + Bb + Cc + \ldots,$$

но он не требует, чтобы коэффициенты $A, B, C,\ldots$ этого выражения были корнями из единицы, а берет за $A, B, C,\ldots$ некоторые целые рациональные числа, такие, чтобы были численно различны все $n! = 1\cdot 2\cdot 3\ldots n$ значений $V, V', V'',\ldots, V^{(n!-1)}$, которые получаются, если в V переставить корни $a, b, c,\ldots$ всеми $n!$ возможными способами. Это всегда можно сделать. Далее, Галуа составляет то уравнение $n!$-й степени, корнями которого являются $V, V', V'',\ldots, V^{(n!-1)}$. Нетрудно показать при помощи теоремы о симметрических многочленах, что коэффициенты этого уравнения $\Phi(x) = 0$ степени $n!$ будут рациональными числами.

До сих пор все довольно похоже на то, что делал Лагранж.

Далее Галуа вводит первое важное новое понятие — понятие неприводимости многочлена в данном поле чисел. Если задан некоторый многочлен от x, коэффициенты которого, например, рациональны, то многочлен называется приводимым в поле рациональных чисел, если он может быть представлен в виде произведения многочленов более низких степеней с рациональными коэффициентами. Если нет, то многочлен называется неприводимым в поле рациональных чисел. Многочлен x^3-x^2-4x-6 приводим в поле рациональных чисел, так как он равен $(x^2+2x+2)(x-3)$, а, например, многочлен x^3+3x^2+3x-5, как это можно показать, неприводим в поле рациональных чисел.

Существуют способы, правда, требующие длинных вычислений, для того чтобы разложить любой заданный многочлен с рациональными коэффициентами на неприводимые множители в поле рациональных чисел.

Галуа предлагает разложить полученный им многочлен $\Phi(x)$ на неприводимые множители в поле рациональных чисел.

Пусть $F(x)$ — один из таких неприводимых множителей (какой из них, для дальнейшего все равно) и пусть он m-й степени.

Многочлен $F(x)$ будет тогда произведением m из $n!$ множителей 1-й степени $x-V, x-V', \ldots, x-V^{(n!-1)}$, на которые разлагается многочлен $n!$-й степени $\Phi(x)$. Пусть этими m множителями являются $x-V, x-V', \ldots, x-V^{(m-1)}$. Перенумеруем как-либо числами (номерами) $1, 2, \ldots, n$ корни $a, b, \ldots, l$ заданного уравнения (6) n-й степени. Тогда в $V, V', \ldots, V^{(n!-1)}$ входят все возможные $n!$ перестановок нумеров $1, 2, \ldots, n$ корней, а в $V, V', \ldots, V^{(m-1)}$ — только m из них. Совокупность G этих m перестановок номеров $1, 2, \ldots, n$ называется группой Галуа заданного уравнения (6)[1].

---

[1] Более подробно о группе Галуа будет идти речь в § 5 главы XX (том 3).

Далее Галуа вводит еще некоторые новые понятия и проводит хотя и простые, но поистине замечательные рассуждения, из которых получается, что условие, необходимое и достаточное для того, чтобы уравнение (6) решалось в радикалах, заключается в том, чтобы группа G перестановок номеров $1, 2, \ldots, n$ удовлетворяла некоторому определенному условию.

Таким образом, предвидение Лагранжа, что в основе всего вопроса лежит теория перестановок, оказалось правильным.

В частности, теорема Абеля о неразрешимости общего уравнения 5-й степени в радикалах может быть теперь доказана так. Можно показать, что существует сколько угодно уравнений 5-й степени, даже с целыми рациональными коэффициентами, таких, для которых соответственный многочлен 120-й степени $\Phi(x)$ неприводим, т. е. таких, группа Галуа которых есть группа всех $5! = 120$ перестановок номеров 1, 2, 3, 4, 5 их корней. Но группа эта, как это можно доказать, не удовлетворяет критерию (признаку) Галуа, и поэтому такие уравнения 5-й степени не решаются в радикалах.

Так, например, можно показать, что уравнение $x^5 + x - a = 0$, где a — положительное целое число, большей частью не решается в радикалах. Например, оно не решается в радикалах при $a = 3, 4, 5, 7, 8, 9, 10, 11\ldots$

Приложение теории Галуа к вопросу о разрешимости геометрической задачи циркулем и линейкой. Одно из замечательных частных приложений теории Галуа следующее. Многие задачи планиметрии могут быть решены построениями циркулем и линейкой, а другие нет. Например, можно построить циркулем и линейкой правильный треугольник, квадрат, пятиугольник, шестиугольник, восьмиугольник, десятиугольник и т. д., но нельзя построить правильного семиугольника, девятиугольника, одиннадцатиугольника и т. д. Какие задачи можно решить циркулем и линейкой, а какие нет? Это было до Галуа неразрешимым вопросом. Из теории Галуа получается следующее его решение.

Совместное решение уравнений двух прямых, прямой и окружности или двух окружностей сводится к решению уравнений первой или второй степени. Для прямой и окружности это очевидно, для случая же двух окружностей $(x - a_1)^2 + (y - b_1)^2 = r_1^2$ и $(x - a_2)^2 + (y - b_2)^2 = r_2^2$, если вычесть одно уравнение из другого, то x^2 и y^2 взаимно уничтожаются и получается уравнение 1-й степени, которое уже надо решать с уравнением одной из окружностей, что тоже дает квадратное уравнение. Поэтому каждый этап задачи, решаемой циркулем и линейкой, сводится к уравнению 1-й или 2-й степени, а вся задача, следовательно, — к алгебраическому уравнению с одним неизвестным, решение которого приводится к цепи квадратных радикалов. Наоборот, если решение какой-либо геометрической задачи сводится к такому алгебраическому уравнению, то она может быть решена циркулем и линейкой, так как квадратный радикал, как известно, может быть построен циркулем и линейкой.

Если дана какая-нибудь геометрическая задача, надо составить то алгебраическое уравнение, к решению которого она сводится. Если такого алгебраического уравнения составить нельзя, задача заведомо не решается циркулем и линейкой. Если оно найдено, надо выделить тот его неприводимый множитель, который связан с решением задачи, и узнать, решается ли это неприводимое уравнение в квадратных радикалах. Как показывает теория Галуа, для этого необходимо и достаточно, чтобы число m перестановок, из которых состоит его группа Галуа, было степенью двойки.

При помощи этого признака оказалась доказанной теорема, высказанная Гауссом о том, что правильный многоугольник с простым числом сторон p может быть построен циркулем и линейкой тогда и только тогда, когда простое число p имеет вид $2^{2^n}+1$, т. е. для $p=3, 5, 17, 257$ и т. д. можно, а для $p=7, 11, 13, 19, 23, 29, 31$ и т. д. нельзя построить правильного многоугольника циркулем и линейкой. Гаусс доказал только первое из этих утверждений.

Этим же способом доказывается, что нельзя циркулем и линейкой разделить произвольный угол на три равные части, нельзя решить задачу удвоения куба, т. е. по ребру куба найти ребро куба вдвое большего объёма, и т. д.

Невозможность квадратуры круга, т. е. невозможность, пользуясь построениями циркулем и линейкой, по радиусу круга найти сторону равновеликого ему квадрата, доказывается иначе. А именно: можно доказать, что сторона такого квадрата не связана с радиусом никаким алгебраическим уравнением, т. е. что она, как говорят, трансцендентна по отношению к радиусу, и, следовательно, подавно не выражается через радиус цепью квадратных радикалов. Доказательство это трудное и не следует из теории Галуа.

Две основные нерешенные задачи, связанные с теорией Галуа. В теории Галуа остались два основных дальнейших вопроса, которые не удалось разрешить в общем виде и до настоящего времени, хотя над ними неустанно работали многие крупные математики.

Первый — это вопрос о так называемых резольвентах Гильберта — Чеботарева (не путать с резольвентами Лагранжа), являющийся прямым обобщением вопроса о решении уравнения в радикалах. Дело в том, что сказать, что уравнение решается в радикалах, или сказать, что его решение сводится к решению цепи последовательных двучленных уравнений,— это одно и то же, так как радикал $\sqrt[n]{A}$ есть, иначе говоря, корень двучленного уравнения $x^n=A$. Но может быть, что хотя уравнение и нельзя свести к цепи таких простых уравнений как двучленные, его все же можно свести к цепи каких-либо других очень простых уравнений. Еще в конце XVIII в. было показано, что общее уравнение 5-й степени можно свести к цепи двучленных уравнений и еще одному уравнению вида $x^5+x+A=0$, которое, хотя и не двучленное, но имеет только один параметр A, т. е. как и двучленные,— однопараметрическое.

Позже было показано, что уже уравнение 6-й степени нельзя свести к цепи однопараметрических уравнений. Требуется для уравнений каждой степени решить вопрос, к цепи каких простейших, т. е. с минимальными числами параметров, уравнений оно может быть сведено.

Если заданное уравнение сводится к цепи однопараметрических уравнений определенных типов, то для каждого из этих однопараметрических уравнений можно вычислить таблицу, дающую его корень по величине его параметра. Тогда решение заданного уравнения сводится к использованию цепи таких таблиц.

Второй, еще более глубокий вопрос состоит в обращении теории Галуа. Галуа показал, что свойства решения уравнения зависят от его группы. Но обратно, всякая ли группа перестановок может быть группой Галуа некоторого уравнения и каковы все те уравнения, группой которых она является.

Относительно первого вопроса были известны лишь частные результаты, хотя над ними упорно работали такие крупнейшие математики, как Клейн и Гильберт; первые общие теоремы были доказаны замечательным советским алгебраистом Н. Г. Чеботаревым.

Второй вопрос для так называемых разрешимых групп, т. е. групп, удовлетворяющих критерию Галуа (см. стр. 264), решен в положительном смысле в последние годы советским математиком И. Р. Шафаревичем.

§ 3. ОСНОВНАЯ ТЕОРЕМА АЛГЕБРЫ

В предыдущем параграфе мы рассмотрели попытки, длившиеся три столетия, дать решение в радикалах для уравнения n-й степени. Вопрос оказался очень глубоким и трудным и привел к созданию новых идей, важных не только для алгебры, но и для всей математики в целом. Что же касается практического решения уравнений, то итогом всей этой огромной работы было следующее. Выяснилось, что решение в радикалах есть далеко не у всех алгебраических уравнений, а если оно и есть, то, в силу своей сложности, за исключением случая квадратного уравнения, оно мало пригодно для практики.

Ввиду этого математики давно стали работать над теорией алгебраических уравнений еще в трех совсем других направлениях, а именно: 1) над вопросом о существовании корня, 2) над вопросом о том, как по коэффициентам уравнения, не решая его, узнать что-нибудь о его корнях, например имеет ли оно действительные корни и сколько их; наконец, 3) над приближенным вычислением корней уравнения.

Прежде всего надо было доказать, что вообще всякое алгебраическое уравнение n-й степени с действительными или комплексными коэффициентами всегда имеет хотя бы один действительный или комплексный корень[1].

---

[1] Дело в том, что существуют неалгебраические уравнения, например $a^x = 0$, которые совсем не имеют корней ни действительных, ни комплексных.

Теорему эту, являющуюся одной из самых важных теорем всей математики, долго не удавалось строго доказать. В силу фундаментальности и трудности доказательства ее называют обычно «основной теоремой алгебры», хотя, по существу дела, большинство ее доказательств относится по своему методу столько же к алгебре, сколько и к анализу бесконечно малых. Первое доказательство ее было дано Даламбером. В одном пункте доказательство Даламбера, как после выяснилось, оказалось недостаточным. А именно, Даламбер принимал за очевидную ту общую лемму анализа, что непрерывная функция, заданная на ограниченном и замкнутом множестве точек, имеет на нем где-нибудь минимум. Это верно, но это надо было доказать. Строгое доказательство этого обстоятельства было получено только во второй половине XVIII в., т. е. через сто лет после исследований Даламбера.

Обычно считается, что первые строгие доказательства основной теоремы алгебры были даны Гауссом; однако некоторые из его доказательств для полной строгости требуют дополнений не меньших, чем их требует доказательство Даламбера. Сейчас известен ряд различных вполне строгих доказательств этой теоремы.

В настоящем параграфе мы рассмотрим доказательство основной теоремы алгебры, основанное на лемме Даламбера, причем приведем полное доказательство и упомянутой леммы анализа.

Теория комплексных чисел. Перед рассмотрением доказательств основной теоремы алгебры прежде всего надо напомнить изучаемую еще в средней школе теорию комплексных чисел. Впервые те трудности, которые привели к созданию теории комплексных чисел, встретились уже при решении квадратного уравнения. Что делать, если число $\frac{p^2}{4} - q$, стоящее под корнем квадратным в формуле решения квадратного уравнения, отрицательно? Нет никакого действительного числа, ни положительного, ни отрицательного, которое является корнем квадратным из отрицательного числа, так как квадрат любого действительного числа есть число положительное или нуль.

После долгих сомнений, длившихся более столетия, математики пришли к заключению, что надо ввести новый вид чисел, так называемые комплексные числа, со следующими правилами действий над ними.

Условно вводится число новой природы $i = \sqrt{-1}$, такое, что $i^2 = -1$, и рассматриваются числа вида $a + bi$, где a и b — обыкновенные действительные числа. Числа $a + bi$ называются комплексными. Два таких числа $a + bi$ и $c + di$ считаются равными, если $a = c$, $b = d$. Суммой двух таких чисел считается число $(a + c) + (b + d)i$, а разностью — число $(a - c) + (b - d)i$. При умножении уславливаются множить эти числа как двучлены, но принимать во внимание, что $i^2 = -1$, т. е.

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

Если a и b считать прямоугольными координатами точки, а точку эту сопоставлять комплексному числу $a + bi$, то указанному сложению и вычитанию комплексных чисел, таким образом, соответствует сложение и вычитание векторов (направленных отрезков), идущих из начала к соответственным точкам с координатами (a, b) и (c, d), так как при сложении векторов как раз складываются соответственные их координаты Какой геометрический смысл имеет на этой так называемой плоскости комплексных чисел вышеописанное умножение, проще всего видно, если рассматривать длину ρ вектора, идущего из начала координат в точку (x, y), (эту длину называют *модулем* комплексного числа $z = x + iy$) и угол φ, который этот вектор образует с осью Ox (этот угол называют *аргументом* комплексного числа $z = x + iy$), другими словами, если рассматривать не декартовы координаты x и y точки, соответствующей комплексному числу z, а ее так называемые полярные координаты ρ и φ (рис. 1). Тогда $x = \rho\cos\varphi$, $y = \rho\sin\varphi$, и, следовательно, самое комплексное число запишется так:

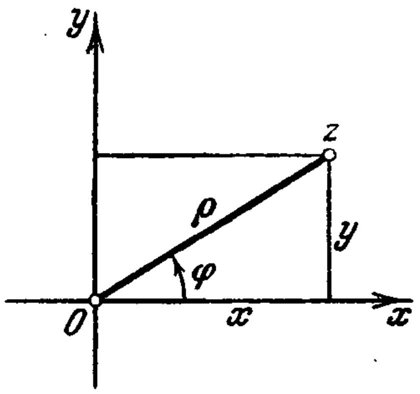


Рис. 1.

$$x + iy = \rho(\cos\varphi + i\sin\varphi).$$

Если
$$a + bi = \rho_1(\cos\varphi_1 + i\sin\varphi_1), \quad c + di = \rho_2(\cos\varphi_2 + i\sin\varphi_2),$$
то
$$ac - bd = \rho_1\rho_2(\cos\varphi_1\cos\varphi_2 - \sin\varphi_1\sin\varphi_2) = \rho_1\rho_2\cos(\varphi_1 + \varphi_2),$$
$$bc + ad = \rho_1\rho_2(\sin\varphi_1\cos\varphi_2 + \cos\varphi_1\sin\varphi_2) = \rho_1\rho_2\sin(\varphi_1 + \varphi_2),$$

откуда мы видим, что при умножении двух комплексных чисел модули их ρ_1 и ρ_2 перемножаются, а аргументы φ_1 и φ_2 складываются. При делении, так как оно есть действие, обратное умножению, наоборот, модули делятся один на другой, а аргументы вычитаются

$$\rho_1(\cos\varphi_1 + i\sin\varphi_1)\rho_2(\cos\varphi_2 + i\sin\varphi_2) = \rho_1\rho_2[\cos(\varphi_1 + \varphi_2) + i\sin(\varphi_1 + \varphi_2)]$$
и
$$\frac{\rho_1(\cos\varphi_1 + i\sin\varphi_1)}{\rho_2(\cos\varphi_2 + i\sin\varphi_2)} = \frac{\rho_1}{\rho_2}[\cos(\varphi_1 - \varphi_2) + i\sin(\varphi_1 - \varphi_2)].$$

При возвышении в степень с целым положительным показателем n, следовательно, модуль возвышается в ту же n-ю степень, а аргумент умножается на n
$$[\rho(\cos\varphi + i\sin\varphi)]^n = \rho^n(\cos n\varphi + i\sin n\varphi).$$

При извлечении корня обратно
$$\sqrt[n]{\rho(\cos\varphi + i\sin\varphi)} = \sqrt[n]{\rho}\left(\cos\frac{\varphi}{n} + i\sin\frac{\varphi}{n}\right).$$

Однако при извлечении корня имеет место еще одно специальное обстоятельство. Пусть n — целый положительный показатель. Тогда

$$\sqrt[n]{\rho(\cos\varphi + i\sin\varphi)}$$

равен числу

$$\sqrt[n]{\rho}\left(\cos\frac{\varphi}{n} + i\sin\frac{\varphi}{n}\right),$$

так как возвышение этого числа в n-ю степень дает подкоренное число.

Но это только одно значение корня. Дело в том, что комплексное число

$$\sqrt[n]{\rho}\left[\cos\left(\frac{\varphi}{n} + \frac{2k\pi}{n}\right) + i\sin\left(\frac{\varphi}{n} + \frac{2k\pi}{n}\right)\right],$$

где k — любое из чисел $1, 2, \ldots, n-1$, также будет корнем n-й степени из числа

$$\rho(\cos\varphi + i\sin\varphi).$$

Действительно, по правилу возвышения в n-ю степень при возвышении такого числа в n-ю степень получается число

$$\left(\sqrt[n]{\rho}\right)^n\left[\cos n\left(\frac{\varphi}{n} + \frac{2k\pi}{n}\right) + i\sin n\left(\frac{\varphi}{n} + \frac{2k\pi}{n}\right)\right] =$$
$$= \rho\left[\cos(\varphi + 2k\pi) + i\sin(\varphi + 2k\pi)\right],$$

причем слагаемые $2k\pi$ под знаком косинуса и синуса можно здесь не писать, так как от этого ни косинус, ни синус не изменятся. Таким образом, n-я степень этого числа тоже есть число

$$\rho(\cos\varphi + i\sin\varphi),$$

т. е. это число есть

$$\sqrt[n]{\rho(\cos\varphi + i\sin\varphi)}.$$

Легко видеть, что никакое другое комплексное число, кроме этих n чисел (для $k = 0, 1, 2, \ldots, n-1$), уже не есть корень n-й степени из

$$\rho(\cos\varphi + i\sin\varphi).$$

Геометрически извлечение корня n-й степени обозначает следующее. Точки комплексной плоскости, соответствующие значениям $\sqrt[n]{}$ из числа $\rho(\cos\varphi + i\sin\varphi)$, лежат в вершинах правильного n-угольника, вписанного в окружность, описанную вокруг начала радиусом $\sqrt[n]{\rho}$, и так повернутого, что одна из вершин этого n-угольника имеет аргумент $\frac{\varphi}{n}$ (рис. 2).

Сделаем следующее замечание. Если

$$f(z) = z^n + c_1 z^{n-1} + \ldots + c_{n-1}z + c_n$$

— многочлен от z с заданными действительными или комплексными коэффициентами $c_1, c_2, \ldots, c_n$, и мы будем изменять z непрерывно, т. е. непрерывно передвигать точку $z = x + iy$ по комплексной плоскости, тогда и комплексная точка $Z = X + iY = f(z)$ также будет передвигаться по комплексной плоскости непрерывно. Это ясно из того, что если подставить в $f(z)$ значения $z = x + iy$, $c_1 = a_1 + b_1 i$, $c_2 = a_2 + b_2 i$, ..., $c_n = a_n + b_n i$ и произвести все вычисления, то получится, что

$$f(z) = X + iY,$$

где

$$X = P(x, y), \qquad Y = Q(x, y)$$

— многочлены n-й степени от x и y с действительными коэффициентами, выражающимися через все a_i и b_i. При непрерывном изменении x и y эти многочлены также будут изменяться непрерывно.

Заметим еще, что так как модуль $\rho = |f(z)|$ равен $\sqrt{X^2 + Y^2}$, то при непрерывном передвижении точки z по комплексной плоскости и модуль $|f(z)|$ будет также изменяться непрерывно. Другими словами, если точка z достаточно близка к точке α, то разность $|f(z)| - |f(\alpha)|$ по абсолютной величине меньше любого наперед заданного положительного числа.

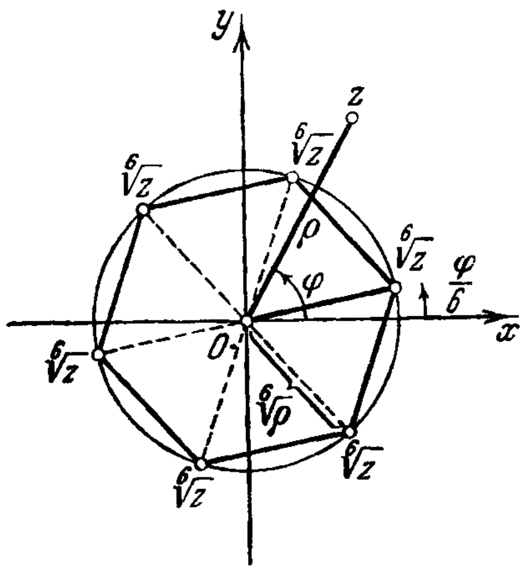


Рис. 2.

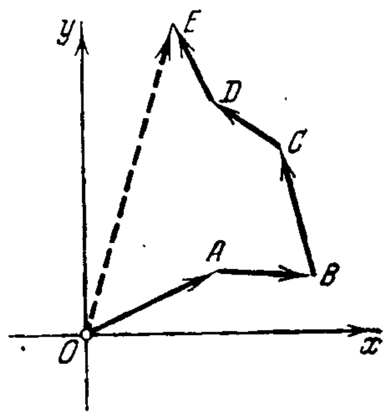


Рис. 3.

Отметим еще, что модуль суммы нескольких комплексных чисел всегда меньше или равен сумме модулей этих чисел, что эквивалентно тому, что замыкающий прямолинейный отрезок OE (рис. 3) короче ломаной $OABCDE$ и равен ей тогда и только тогда, когда все ее отрезки лежат на одной прямой и идут в одну сторону.

Напомним, наконец, что сказать «комплексное число равно нулю» или «его модуль равен нулю» — одно и то же, так как модуль ρ комплексного числа есть расстояние соответствующей ему точки от точки нуль.

Мы сейчас применим теорию комплексных чисел к доказательству основной теоремы алгебры, однако значение теории комплексных чисел далеко выходит за пределы алгебры. Во многих других частях математики, как и в алгебре, нельзя обойтись без них. Во многих приложениях, например в теории переменных токов, ряд вопросов наиболее просто решается при помощи комплексных чисел. Но что наиболее важно — это применение комплексных чисел, собственно теории функций от комплексного переменного, к теории некоторых специальных функций от двух действительных переменных, которые называются гармоническими. При помощи этих функций решают важные вопросы теории полета самолета, теории движения теплоты в пластинке, теории плоского электрического поля и некоторые вопросы теории упругости. Знаменитая теорема о поддерживающей силе крыла самолета была получена создателем современной аэродинамики Н. Е. Жуковским при помощи исследования функций комплексного переменного [1].

Перейдем теперь к доказательству основной теоремы алгебры.

Теорема. Любой многочлен

$$f(z) = a_0 z^n + a_1 z^{n-1} + \ldots + a_{n-1} z + a_n,$$

коэффициенты которого

$$a_0,\ a_1,\ \ldots,\ a_{n-1},\ a_n$$

— любые заданные действительные или комплексные числа, имеет хотя бы один, действительный или комплексный, корень.

Мы будем предполагать, что заданный многочлен n-й степени, т. е. что $a_0 \neq 0$.

Поверхность модуля многочлена. Рассмотрим весь вопрос геометрически. Будем над каждой точкой z комплексной плоскости восставлять аппликату (перпендикуляр) t, равную по своей длине модулю $|f(z)|$ многочлена $f(z)$ в этой точке z. Концы таких аппликат образуют некоторую поверхность M, которую можно назвать поверхностью модуля многочлена $f(z)$. Мы видим, что эта поверхность: 1) нигде не опускается под комплексную плоскость, так как модуль любого комплексного числа [в данном случае числа $f(z)$] не отрицателен; 2) для любой точки z комплексной плоскости поверхность эта имеет одну и только одну точку, которая лежит вертикально — над этой точкой или в самой этой точке, т. е. поверхность M простирается одним листом над всей комплексной плоскостью и, может быть, в некоторых точках доходит до самой этой плоскости; 3) поверхность эта непрерывна в том смысле, что при непрерывном изменении положения точки z на комплексной плоскости непрерывно изменяется величина $t = |f(z)|$, т. е. аппликата t точек этой поверхности (это было доказано на стр. 270).

Основная теорема алгебры состоит в том, чтобы доказать, что поверхность M хотя бы в одной точке опускается до самой комплексной плоскости, а не везде проходит на некоторой высоте над нею.

О возрастании модуля многочлена при удалении от начала. Покажем, что какое бы большое положительное число G ни задать, можно указать такой радиус R, что для всех точек z комплексной плоскости, лежащих вне круга радиуса R с центром в начале, аппликаты t точек поверхности M больше G.

[1] См. главу IX (том 2).

Действительно, напишем многочлен $f(z)$ так:

$$a_0z^n\left[1+\left(\frac{a_1}{a_0z}+\frac{a_2}{a_0z^2}+\cdots+\frac{a_n}{a_0\,z^n}\right)\right].$$

Модуль выражения $\left(\frac{a_1}{a_0z}+\frac{a_2}{a_0z^2}+\cdots+\frac{a_n}{a_0z^n}\right)$ не больше суммы модулей слагаемых $\left|\frac{a_1}{a_0z}\right|+\left|\frac{a_2}{a_0z^2}\right|+\cdots+\left|\frac{a_n}{a_0z^n}\right|$, причем при увеличении модуля z каждое из слагаемых этой суммы уменьшается, а следовательно, уменьшается и эта сумма. Поэтому для всех z, по модулю бо́льших некоторого числа R', модуль этой скобки меньше, например, чем $^1/_2$.

Но тогда для всех таких z скобка $\Omega=\left[1+\left(\frac{a_1}{a_0z}+\frac{a_2}{a_0z^2}+\cdots+\frac{a_n}{a_0z^n}\right)\right]$ будет по модулю больше $^1/_2$. Модуль первого множителя a_0z^n равен $|a_0|\cdot|z|^n$, и, следовательно, с увеличением модуля z возрастает, и притом сколь угодно возрастает. Поэтому какое бы большое положительное число G ни задать, существует такое положительное число R, что для всех z, по модулю больших R, уже $|f(z)|=|a_0|\cdot|z|^n\cdot|\Omega|$ больше, чем G.

Существование минимумов поверхности M. Мы будем говорить, что в точке α комплексной плоскости имеется минимум поверхности M, если значение апликаты t точки поверхности M в этой точке α меньше или равно ее значениям во всех точках некоторой окрестности точки α, т. е. во всех точках некоторого, хотя бы очень маленького, кружка с центром в точке α.

Пусть апликата t точки поверхности M, соответствующая началу, т. е. точке $z=0$, комплексной плоскости, равна g, т. е. $|f(0)|=g$. Возьмем $G>g$. Все апликаты точек t поверхности M не отрицательны и непрерывно изменяются при непрерывном передвижении точки z по комплексной плоскости. Поверхность M вне круга, описанного вокруг начала радиусом R, имеет апликаты $t>G$, а в его центре имеет апликату $t=g<G$. Даламбер считал очевидным следствием из этого, что где-нибудь внутри круга R есть такая точка, в которой апликата наименьшая, точнее, в которой апликата t точки поверхности M меньше или равна апликатам всех остальных точек круга R, т. е. что поверхность M имеет хотя бы один минимум.

Строгое доказательство существования такого минимума основывается на следующей *аксиоме непрерывности* совокупности действительных чисел.

Если заданы две последовательности действительных чисел: $a_1\leqslant a_2\leqslant\ldots\leqslant a_n\leqslant\ldots$ и $b_1\geqslant b_2\geqslant\ldots\geqslant b_n\geqslant\ldots$, такие, что $b_n>a_n$ при всех n и $b_n-a_n\to 0$ при $n\to\infty$, то существует одно и только одно действительное число c, такое, что $a_n\leqslant c\leqslant b_n$ при всех n.

Геометрически свойство непрерывности означает, что если на прямой задана последовательность отрезков $[a_n, b_n]$ (рис. 4), таких, что каждый последующий содержится в предыдущем, и длины отрезков становятся сколь угодно малыми, то существует точка c, принадлежащая всем отрезкам последовательности. Иными словами, отрезки «стягиваются» к некоторой точке, а не к «пустому месту».

В силу того, что длина отрезка $[a_n, b_n]$ при увеличении n стремится к нулю, такая точка c только одна. Из этого свойства непрерывности для совокупности действительных чисел, т. е. для совокупности всех точек на числовой оси, немедленно следует свойство непрерывности для комплексных чисел, т. е. для точек плоскости. Мы дадим геометрическую формулировку этого свойства.

Если на плоскости дана последовательность прямоугольников $\Delta_1, \Delta_2, \ldots, \Delta_n, \ldots$ со сторонами, параллельными координатным осям, из которых каждый последующий содержится в предыдущем и диагонали которых неограниченно уменьшаются, то

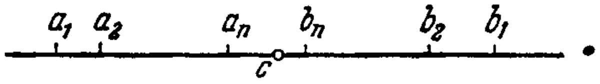


Рис. 4.

существует одна и только одна точка, принадлежащая всем прямоугольникам последовательности. Это свойство непрерывности плоскости непосредственно следует из свойства непрерывности для прямой. Для доказательства достаточно спроектировать прямоугольники на координатные оси.

Теперь легко установить так называемую теорему Больцано — Вейерштрасса.

Если в некотором прямоугольнике задана бесконечная последовательность точек $z_1, z_2, \ldots, z_n, \ldots$, то внутри или на границе прямоугольника найдется такая точка z_0, в любой сколь угодно малой окрестности которой (т. е. внутри любого достаточно малого кружка с центром в точке z_0) существует бесконечно много точек из последовательности $z_1, z_2, \ldots, z_n, \ldots$

Для доказательства обозначим заданный прямоугольник через Δ_1. Разобьем его на четыре равные части прямыми, параллельными осям координат. Хотя бы в одной из частей должно оказаться бесконечно много точек данной последовательности. Такую часть обозначим через Δ_2. Прямоугольник Δ_2 снова разобьем на четыре равные части и выберем из них ту Δ_3, в которой содержится бесконечно много точек данной последовательности, и т. д.

Получим последовательность вложенных прямоугольников $\Delta_1, \Delta_2, \Delta_3, \ldots$, диагонали которых безгранично убывают. По свойству непрерывности найдется точка z_0, принадлежащая всем построенным прямоугольникам. Она и будет искомой. Действительно, какую бы малую окрестность точки z_0 ни взять, прямоугольники последовательности $\Delta_1, \Delta_2, \ldots, \Delta_n, \ldots$, начиная с некоторого, окажутся внутри этой окрестности, как только диагональ станет меньше радиуса окрестности, а каждый из прямоугольников содержит бесконечно много точек из последовательности $z_1, z_2, \ldots, z_n, \ldots$ Теорема Больцано — Вейерштрасса доказана.

Теперь уже легко доказать теорему о минимуме модуля $|f(z)|$ многочлена. Пусть, как и раньше, $|f(0)| = g$; G — некоторое число, большее g; R такое, что, при $|z| > R$, $|f(z)| > G$.

Если $g = 0$, т. е. $f(0) = 0$, то модуль $|f(z)|$ многочлена имеет минимум в точке 0, так как во всех точках он $\geqslant 0$.

Если $g > 0$ и $|f(z)| \geqslant g$ для всех точек z, то $|f(z)|$ тоже имеет в точке 0 минимум. Пусть $g > 0$ и существуют точки z, в которых $|f(z)| < g$; тогда в ряду чисел

$$0, \ \frac{g}{n}, \ \frac{2g}{n}, \ldots, \ \frac{ng}{n} = g \qquad (*)$$

найдем наибольшее $c_n = \frac{i}{n} g$, такое, что все значения $|f(z)| \geqslant c_n$. Для следующего числа $c'_n = \frac{i+1}{n} g$ ряда $(*)$ найдется хотя бы одна точка z_n, такая, что $|f(z_n)| < c'_n$.

Будем бесконечно увеличивать n. Для всех n имеет место $|z_n| \leqslant R$, ибо если $|z_n| > R$, то $|f(z_n)|$ был бы больше G, а следовательно, и больше g.

Таким образом, все точки z_n лежат внутри прямоугольника со стороной $2R$ и с центром в начале координат. Быть может, некоторые из этих точек совпадают друг с другом.

По теореме Больцано — Вейерштрасса найдется точка z_0, в любой окрестности которой существует бесконечно много точек из последовательности $z_1, z_2, \ldots, z_n, \ldots$

Установим, что точка z_0 как раз и доставляет искомый минимум $|f(z)|$.

Действительно, пусть z любая точка. Тогда

$$|f(z)| \geqslant c_n = c'_n - \frac{g}{n} > |f(z_n)| - \frac{g}{n} = |f(z_0)| + (|f(z_n)| - |f(z_0)|) - \frac{g}{n}.$$

Это неравенство верно при любом n. Если взять для n такие значения, чтобы z_n неограниченно приближалось к z_0, то, вследствие непрерывности $|f(z)|$, $|f(z_n)| - |f(z_0)|$ станет сколь угодно малым по абсолютной величине так же, как и $\frac{g}{n}$.

Следовательно, $|f(z)| \geqslant |f(z_0)|$, т. е. $|f(z)|$ действительно достигает минимума в точке z_0.

Лемма Даламбера. Ввиду того, что все аппликаты t точек поверхности M, как модули, неотрицательны, то очевидно, что всякому корню многочлена $f(z)$, т. е. точке z комплексной плоскости, в которой сам многочлен $f(z)$, а следовательно, и его модуль $|f(z)|$, равен нулю, соответствует минимум поверхности модуля M. Однако, как показал Даламбер, верно и обратное: во всяком минимуме поверхность M доходит до самой комплексной плоскости, и, следовательно, в нем имеется корень многочлена $f(z)$. Другими словами, нет минимумов поверхности M, в которых аппликата t положительна, а не нуль. Это следует из так называемой леммы Даламбера:

Если α — какое угодно заданное комплексное число, такое, что $f(\alpha) \neq 0$, то можно всегда найти сколь угодно малое по модулю комплексное число h, такое, что $|f(\alpha + h)| < |f(\alpha)|$.

Доказательство. Рассмотрим многочлен

$$f(\alpha + h) = a_0(\alpha + h)^n + a_1(\alpha + h)^{n-1} + \ldots + a_{n-1}(\alpha + h) + a_n$$

и расположим этот многочлен от двух букв α и h по восходящим степеням h. В этом многочлене будет член, вовсе не содержащий h, а именно

$$a_0\alpha^n + a_1\alpha^{n-1} + \ldots + a_{n-1}\alpha + a_n = f(\alpha) \neq 0,$$

так как было предположено, что $f(\alpha) \neq 0$. Будет также член с h^n, а именно: a_0h^n, так как было предположено, что $a_0 \neq 0$. Что же касается членов с промежуточными степенями h, то некоторые из них, а иногда и все, могут отсутствовать. Пусть наинизшая степень h, которая встречается в этом разложении, есть m-я, где $1 \leqslant m \leqslant n$, т. е. разложение это имеет вид

$$f(\alpha + h) = f(\alpha) + Ah^m + Bh^{m+1} + Ch^{m+2} + \ldots + a_0h^n.$$

Напишем его так:

$$f(\alpha + h) = f(\alpha) + Ah^m + Ah^m\left(\frac{B}{A}h + \frac{C}{A}h^2 + \ldots + \frac{a_0}{A}h^{n-m}\right),$$

где $A \neq 0$, а B, C и т. д. могут и равняться и не равняться нулю.

После этой подготовки, грубо говоря, доказательство леммы Даламбера протекает так. За h берется достаточно малое по модулю комплексное число, такое, чтобы длина вектора Ah^m была меньше длины вектора $f(\alpha)$, и с таким аргументом, чтобы направление вектора Ah^m было обратно направлению вектора $f(\alpha)$. Тогда вектор $f(\alpha)+Ah^m$ будет короче, чем вектор $f(\alpha)$. Но для всех h, достаточно малых по модулю, модуль скобки $\left(\frac{B}{A}h+\frac{C}{A}h^2+\ldots+\frac{a_0}{A}h^{n-m}\right)$ сколь угодно мал, например меньше единицы, и, следовательно, длина вектора $\Delta=Ah^m\left(\frac{B}{A}h+\frac{C}{A}h^2+\cdots\right.$ $\left.\cdots+\frac{a_0}{A}h^{n-m}\right)$ короче длины вектора Ah^m, и потому вектор $f(\alpha+h)=$ $=f(\alpha)+Ah^m+\Delta$, как это видно на рис. 5, также короче вектора $f(\alpha)$, даже если бы направление вектора Δ было прямо противоположно направлению вектора Ah^m.

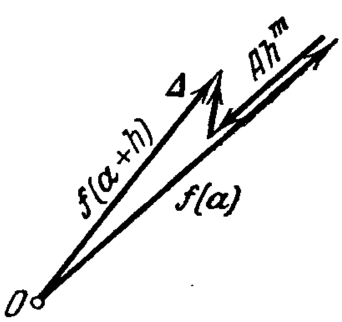


Рис. 5.

Подробности этого доказательства следующие:

1. Так как при умножении аргументы множителей складываются, аргумент h надо взять таким, чтобы

$$\arg A+m\cdot\arg h=\arg f(\alpha)+180^\circ,$$

т. е. надо взять

$$\arg h=\frac{\arg f(\alpha)-\arg A+180^\circ}{m}.$$

2. Модуль скобки $\left(\frac{B}{A}h+\frac{C}{A}h^2+\cdots+\frac{a_0}{A}h^{n-m}\right)$ не больше суммы модулей ее слагаемых $T=\left|\frac{B}{A}h\right|+\left|\frac{C}{A}h^2\right|+\cdots+\left|\frac{a_0}{A}h^{n-m}\right|$, причем при уменьшении модуля h каждое из слагаемых этой суммы сколь угодно уменьшается, следовательно, уменьшается и вся эта сумма. Поэтому, если h — комплексное число написанного выше аргумента и h_0 — такой модуль, что для h, по модулю меньших, чем h_0, выполняются оба условия $|Ah^m|<|f(\alpha)|$ и $T<1$, то для всех h этого аргумента, по модулю меньших h_0, уже будет $|f(\alpha+h)|<|f(\alpha)|$, что и доказывает лемму Даламбера.

Из леммы Даламбера непосредственно следует, что всякий минимум поверхности M модуля многочлена $f(z)$ дает корень этого многочлена. Действительно, если бы в точке α было $f(\alpha)\neq 0$, то в силу леммы Даламбера в сколь угодно близких к ней точках $\alpha+h$ было бы $|f(\alpha+h)|<|f(\alpha)|$, т. е. не существовало бы такого кружка с центром в точке α, во всех точках которого модуль $f(z)$ не меньше модуля $f(\alpha)$, и поэтому в точке α не могло бы быть минимума модуля $f(z)$. Тем самым основная теорема высшей алгебры доказана.

Общий вид поверхности M модуля. Поверхность M модуля многочлена $f(z)$ лежит над комплексной плоскостью z. Она имеет вид, указанный на рис. 6. Можно показать, что на большой высоте t поверхность M весьма мало отличается от поверхности, получающейся от вращения параболы n-го порядка $t = |a_0| x^n$ вокруг оси Ot. Но при малых t поверхность M имеет минимумы, число которых равно числу различных корней уравнения $f(z) = 0$. Во всех этих минимумах поверхность M упирается в самое комплексную плоскость z.

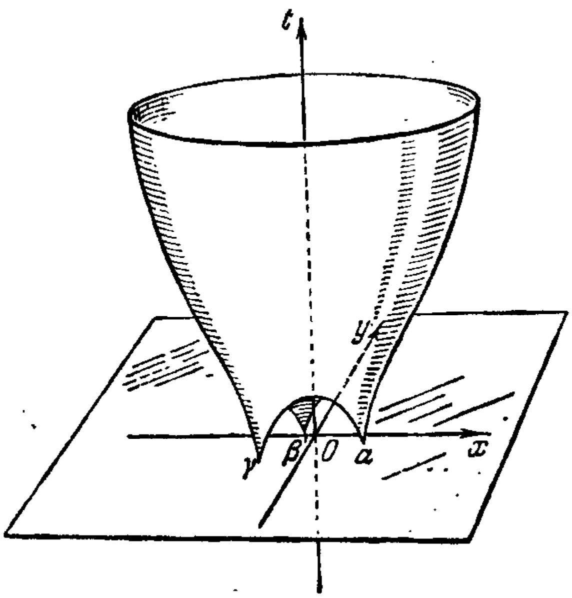


Рис. 6.

§ 4. ИССЛЕДОВАНИЕ РАСПОЛОЖЕНИЯ КОРНЕЙ МНОГОЧЛЕНА НА КОМПЛЕКСНОЙ ПЛОСКОСТИ

Ряд важных для практики вопросов связан с такой задачей: не решая уравнения, получить те или иные сведения относительно расположения его корней на комплексной плоскости. Первым таким вопросом был вопрос об определении числа действительных корней уравнения. То есть, если задано уравнение с действительными коэффициентами, то, не решая его, по какому-нибудь признаку, зависящему от его коэффициентов, сказать, имеет ли оно действительные корни, и если имеет, то сколько; или сколько имеет положительных и сколько отрицательных корней; или сколько оно имеет действительных корней, лежащих между заданными пределами a и b.

Производные от многочлена. В этом параграфе существенную роль будет играть производнан от многочлена. Что такое производная от данной функции, было изложено в главе II.

Для многочлена $a_0x^n + a_1x^{n-1} + \cdots + a_{n-1}x + a_n$ производная равна, как известно, многочлену $na_0x^{n-1} + (n-1)a_1x^{n-2} + \cdots + a_{n-1}$.

Понятие производной в главе II было рассмотрено лишь для функций от действительной переменной. В алгебре необходимо считать переменную принимающей произвольные комплексные значения и вводить в рассмотрение многочлены с комплексными коэффициентами.

Однако определение производной можно сохранить прежнее — как предел отношения приращения функции к приращению независимой переменной. Формула для вычисления производной от многочлена с комплексными коэффициентами, а также основные правила дифференцирования (производная суммы, произведения, степени) остаются прежними[1].

Простые и кратные корни многочлена. В § 2 этой главы было установлено, что если число a — корень многочлена $f(x)$, то $f(x)$ делится на $x-a$ без остатка. Если при этом $f(x)$ не делится на $(x-a)^2$, то число a называется простым корнем многочлена $f(x)$. Вообще, если многочлен $f(x)$ делится на $(x-a)^k$ и не делится на $(x-a)^{k+1}$, то число a называется корнем кратности k.

Корень a кратности k часто рассматривают как k равных корней. Основанием к этому является то, что множитель $(x-a)^k$, присутствующий в разложении $f(x)$ на линейные множители, есть произведение k множителей, равных $(x-a)$.

В силу того, что каждый многочлен степени n разлагается в произведение n линейных множителей, число корней многочлена равно его степени, если условиться считать каждый корень столько раз, какова его кратность.

Верны следующие теоремы:

1. Простой корень многочлена не является корнем его производной.

2. Кратный корень многочлена является корнем его производной на единицу меньшей кратности.

Действительно, пусть $f(x)=(x-a)^k f_1(x)$ и $f_1(x)$ не делится на $(x-a)$, т. е. $f_1(a)\neq 0$. Тогда

$$f'(x)=k(x-a)^{k-1}f_1(x)+(x-a)^k f_1'(x)=$$
$$=(x-a)^{k-1}[kf_1(x)+(x-a)f_1'(x)]=(x-a)^{k-1}F(x).$$

Многочлен $F(x)=kf_1(x)+(x-a)f_1'(x)$ не делится на $(x-a)$, ибо $F(a)=kf_1(a)\neq 0$.

Следовательно, $f'(x)$ при $k=1$ не делится на $x-a$, а при $k>1$ $f'(x)$ делится на $(x-a)^{k-1}$, но не делится на $(x-a)^k$. Тем самым обе теоремы доказаны.

Теорема Ролля и некоторые ее следствия. По известной теореме Ролля[2], если действительные числа a и b являются корнями многочлена с действительными коэффициентами, то существует число c, лежащее между a и b и являющееся корнем производной.

Из теоремы Ролля вытекают интересные следствия:

1. Если все корни многочлена $f(x)=a_0x^n+\cdots+a_n$ действительны, то все корни его производной тоже действительны. При этом между

[1] См. главу IX (том 2).

[2] Эта теорема является простейшим случаем теоремы о среднем, о которой шла речь в главе II на стр. 128.

двумя соседними корнями $f(x)$ существует один корень $f'(x)$, и этот корень простой. Действительно, пусть $x_1 < x_2 < \cdots < x_k$ являются корнями $f(x)$ с кратностями $m_1, m_2, \ldots, m_k$. Ясно, что $m_1 + m_2 + \cdots + m_k = n$.

Тогда производная $f'(x)$ по теореме о кратных корнях имеет корни $x_1, x_2, \ldots, x_k$ с кратностями $m_1 - 1$, $m_2 - 1, \ldots, m_k - 1$ и по теореме Ролля еще по крайней мере по одному корню $y_1, y_2, \ldots, y_{k-1}$ внутри каждого из промежутков $(x_1, x_2), (x_2, x_3), \ldots, (x_{k-1}, x_k)$ между двумя соседними корнями $f(x)$. Таким образом, число действительных корней $f'(x)$ равно (с учетом кратностей) по крайней мере $(m_1 - 1) + (m_2 - 1) + \cdots + (m_k - 1) + k - 1 = n - 1$. Но $f'(x)$ как многочлен $(n-1)$-й степени имеет (с учетом кратностей) $n - 1$ корень. Следовательно, все корни $f'(x)$ действительны, $y_1, y_2, \ldots, y_{k-1}$ являются простыми корнями, и других корней, кроме $x_1, x_2, \ldots, x_k$ и $y_1, y_2, \ldots, y_{k-1}$, многочлен $f'(x)$ не имеет.

2. Если все корни многочлена $f(x)$ действительны и из них p положительных (с учетом кратностей), то $f'(x)$ имеет p или $p - 1$ положительных корней.

Действительно, пусть $x_1 < x_2 < \ldots < x_k$ — все положительные корни многочлена $f(x)$ с кратностями $m_1, m_2, \ldots, m_k$. Тогда $m_1 + m_2 + \cdots + m_k = p$. Производная $f'(x)$ будет иметь следующие положительные корни: $x_1, x_2, \ldots, x_k$ с кратностями $m_1 - 1$, $m_2 - 1, \ldots, m_k - 1$; простые корни $y_1, y_2, \ldots, y_{k-1}$, лежащие в промежутках $(x_1, x_2), \ldots, (x_{k-1}, x_k)$, и, может быть, еще один простой корень y_0, лежащий в промежутке (x_0, x_1), где x_0 — наибольший неположительный корень $f(x)$. Следовательно, число положительных корней $f'(x)$ равно $(m_1 - 1) + \cdots + (m_k - 1) + k - 1 = p - 1$ или $(m_1 - 1) + \cdots + (m_k - 1) + (k - 1) + 1 = p$, что и требовалось доказать.

Правило знаков Декарта. В своей знаменитой книге 1637 г. «Геометрия», в которой было дано первое изложение аналитической геометрии, Декарт между прочим дал и первую замечательную алгебраическую теорему, относящуюся к теории расположения корней многочлена на комплексной плоскости, — так называемое «правило знаков Декарта». Оно может быть высказано так:

Если коэффициенты уравнения действительные и все его корни также заведомо действительные, то число его положительных корней, если учитывать их кратности, равно числу перемен знаков в ряде его коэффициентов. Если же оно имеет и комплексные корни, то число это равно или на некоторое четное число меньше числа этих перемен знаков.

Объясним прежде всего, что такое число перемен знаков в ряде коэффициентов уравнения. Для получения этого числа выписывают все коэффициенты уравнения подряд, например по убывающим степеням неизвестной, включая коэффициент при x^n и постоянный член, но совсем опуская равные нулю коэффициенты, и рассматривают все пары соседних чисел полученного ряда. Если в такой паре знаки чисел различны, то называем это переменой знаков. Например, если дано уравнение

$$x^7 + 3x^5 - 5x^4 - 8x^2 + 7x + 2 = 0,$$

то ряд его коэффициентов

$$1,\ 3,\ -5,\ -8,\ 7,\ 2$$

и перемен знаков в нем две.

Перейдем теперь к доказательству первой части теоремы[1].

Без нарушения общности можно считать, что старший коэффициент a_0 многочлена $f(x) = a_0x^n + \cdots + a_n$ положителен.

Установим прежде всего, что если $f(x)$ имеет только действительные корни и из них p положительных (с учетом кратности), то $(-1)^p$ есть знак последнего, отличного от нуля, коэффициента $f(x)$.

В самом деле, пусть

$$f(x) = a_0x^n + \cdots + a_kx^{n-k} =$$
$$= a_0x^{n-k}(x - x_1)\ldots(x - x_p)(x - x_{p+1})\ldots(x - x_{n-k}),$$

где $x_1, \ldots, x_p$ — положительные корни $f(x)$, $x_{p+1}, \ldots, x_{n-k}$ — отрицательные корни $f(x)$; каждый считается столько раз, какова его кратность. Тогда $a_k = a_0(-1)^p x_1\ldots x_p(-x_{p+1})\ldots(-x_{n+k})$, и в силу положительности чисел $a_0, x_1, \ldots, x_p, -x_{p+1}, -x_{n-k}$ знак a_k есть $(-1)^p$[2].

Дальнейшее доказательство построено по способу математической индукции.

Для многочленов 1-й степени теорема очевидна. Действительно, многочлен 1-й степени $a_0x + a_1$ имеет единственный корень $-\frac{a_1}{a_0}$, который положителен в том и только в том случае, когда a_0 и a_1 имеют противоположные знаки.

Допустим теперь, что теорема доказана для всех многочленов $(n-1)$-й степени с действительными корнями, и в этом предположении докажем ее для многочлена $f(x) = a_0x^n + \cdots + a_{n-1}x + a_n$ степени n.

1. $a_n = 0$. Рассмотрим многочлен $f_1(x) = a_0x^{n-1} + \cdots + a_{n-1}$. Положительные корни многочленов $f(x)$ и $f_1(x)$ одинаковы; число перемен знаков в рядах их коэффициентов тоже одинаково. Для многочлена $f_1(x)$ правило Декарта верно; следовательно, оно верно и для многочлена $f(x)$.

2. $a_n \neq 0$. Введем в рассмотрение производную $f'(x) = na_0x^{n-1} + (n-1)\ a_1x^{n-2} + \cdots + a_{n-1}$.

Очевидно, что число перемен знаков в ряду коэффициентов многочлена $f(x)$ равно аналогичному числу для производной $f'(x)$, если знаки

[1] Может быть дано другое, непосредственное доказательство, не использующее производных, но оно несколько длиннее.

[2] Заметим, что это утверждение справедливо и для случая, когда среди корней $f(x)$ могут быть комплексные.

a_n и последнего, отличного от нуля, коэффициента производной совпадают, или на единицу больше, если эти знаки противоположны.

В силу сказанного в начале доказательства, в первом случае число положительных корней $f(x)$ и $f'(x)$ имеет одинаковую четность, во втором случае — противоположную. Но, как мы вывели из теоремы Ролля, число положительных корней многочлена (если все его корни действительны) может или равняться числу положительных корней его производной, или быть на единицу больше. Принимая это во внимание, заключаем, что в первом случае $f(x)$ имеет столько же положительных корней, сколько $f'(x)$, во втором — на единицу больше. Для $f'(x)$ правило Декарта верно по индуктивному предположению, т. е. число положительных корней $f'(x)$ равно числу перемен знаков в ряду его коэффициентов. Следовательно, в обоих случаях число положительных корней $f(x)$ равно числу перемен знаков в ряду коэффициентов, что и требовалось доказать.

Вторая часть правила Декарта устанавливается не сложнее, но мы не будем приводить ее доказательства.

Замечание 1. Особенно важно первое утверждение теоремы Декарта, так как во многих практических вопросах заведомо известно, что все корни написанного уравнения действительные. В этом случае можно быстро узнать, сколько корней положительных и сколько отрицательных. А сколько у уравнения нулевых корней, видно сразу.

Замечание 2. Если в рассматриваемом многочлене положить $x = y + a$, где a — какое угодно заданное действительное число, т. е. написать многочлен $f(y + a)$, то положительными корнями y этого многочлена будут те и только те, которые получаются из тех корней x заданного многочлена $f(x)$, которые больше, чем a. Поэтому число корней заданного многочлена $f(x)$, все корни которого действительные, лежащих между заданными пределами a и b ($b > a$), равно числу перемен знаков у многочлена $f(y + a)$ минус число перемен знаков у многочлена $f(z + b)$. Если же не все корни $f(x)$ действительные, то можно показать, что оно равно этой разности или на четное число меньше. Это так называемая теорема Бюдана.

Теорема Штурма. Правило знаков Декарта, равно как и теорема Бюдана не дают, однако, ответа на вопросы: имеет ли данное уравнение с действительными коэффициентами хоть один действительный корень, сколько оно имеет всего действительных корней, а тем более сколько оно имеет действительных корней, лежащих между данными пределами a и b. Более двух столетий математики пытались решить эти вопросы, и все безрезультатно. Длинный ряд работ в этом направлении принадлежит Декарту, Ньютону, Бюдану, Сильвестру, Фурье и многим другим, но им не удалось решить даже первого из этих вопросов, пока, наконец, в 1835 г. французский математик Штурм не предложил способа, дающего решение всех трех вопросов.

Вывод способа Штурма собственно совсем не сложен, но он таков, что его можно было бы еще долго искать и все же не найти. Сам Штурм был очень счастлив, что ему удалось решить эту знаменитую и чрезвычайно важную для практики задачу алгебры. Когда он на лекциях доходил до изложения своего результата, он обычно говорил: «Вот теорема, имя которой я ношу». Надо, однако, сказать, что Штурм решил эту задачу не случайно; он много лет думал над смежными с нею вопросами.

Пусть $f(z)$ — многочлен с действительными коэффициентами; $f_1(z)$ — его производная $f'(z)$. Поделим многочлен $f(z)$ на $f_1(z)$, обозначив через $f_2(z)$ остаток при этом делении, взятый с противоположным знаком. Далее, поделим $f_1(z)$ на $f_2(z)$, обозначив остаток, взятый с противоположным знаком, через $f_3(z)$, и т. д.

Можно доказать, что последний, отличный от нуля многочлен $f_1(z)$ построенной последовательности будет некоторым постоянным числом c.

Теорема Штурма состоит в том, что если $a<b$ — два действительных числа, не являющихся корнями многочлена $f(z)$, то, подставив в многочлены

$$f(z),\ f_1(z),\ \ldots,\ f_{s-1}(z),\ c$$

$z=a$ и $z=b$, мы получаем два ряда действительных чисел

$$f(a),\ \ f_1(a),\ \ f_2(a),\ \ \ldots,\ \ f_{s-1},(a),\ c, \tag{I}$$

$$f(b),\ \ f_1(b),\ \ f_2(b),\ \ \ldots,\ \ f_{s-1}(b),\ c, \tag{II}$$

такие, что число перемен знаков в ряде (I) больше или равно числу перемен знаков в ряде (II) и разность между этими числами перемен знаков в точности равна числу действительных корней $f(z)$, лежащих между a и b, или, как иногда говорят, их число равно потере перемен знаков в ряде (I) при переходе от a к b.

Доказательство теоремы Штурма не труднее доказательства теоремы Декарта, однако мы его приводить не будем.

Теорема Штурма дает возможность подсчитать число корней многочлена с действительными коэффициентами на любом отрезке действительной оси. Поэтому применение теоремы Штурма к любому данному многочлену дает возможность установить довольно ясную картину расположения корней многочлена на действительной оси, в частности *отделить* корни, т. е. найти такие промежутки, в каждом из которых содержится по одному корню многочлена.

Во многих приложениях не меньшее значение имеет решение аналогичной задачи для комплексных корней многочлена. Так как комплексные числа изображаются точками не на прямой, а на плоскости, то нельзя говорить о «промежутке», в котором заключается комплексный корень; вместо промежутка приходится рассматривать область, т. е. часть плоскости, выделенную тем или другим способом.

Таким образом, в применении к комплексным корням ставится следующая задача:

Дан многочлен $f(z)$ и дана область на комплексной плоскости. Требуется узнать число корней многочлена внутри этой области.

Будем предполагать, что область ограничена замкнутым контуром (рис. 7) и что на контуре области многочлен $f(z)$ корней не имеет.

Представим себе, что точка z обходит контур области один раз в положительном направлении. Каждое значение многочлена может быть тоже

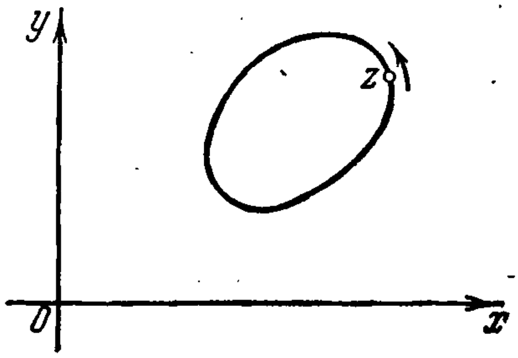


Рис. 7.

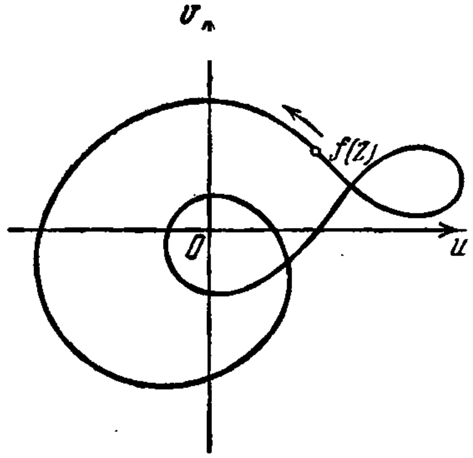


Рис. 8.

изображено точкой на плоскости. При непрерывном изменении z многочлен $f(z)$ тоже изменяется непрерывно. Поэтому, пока z обойдет один раз контур области, $f(z)$ зачертит некоторую замкнутую линию. Эта линия не будет проходить через начало координат, ибо $f(z)$, по предположению, не обращается в нуль ни в одной из точек контура (рис. 8).

Ответ на поставленный выше вопрос дается следующей теоремой:

Принцип аргумента. Число корней многочлена $f(z)$ внутри области, ограниченной замкнутым контуром C, равно числу обходов вокруг начала координат точкой $f(z)$, когда z обходит контур C один раз в положительном направлении.

Для доказательства разложим $f(z)$ на линейные множители

$$f(z) = a_0 z^n + a_1 z^{n-1} + \cdots + a_n = a_0 (z - z_1)(z - z_2) \ldots (z - z_n).$$

Мы знаем, что аргумент произведения нескольких комплексных чисел равен сумме аргументов сомножителей. Следовательно,

$$\arg f(z) = \arg a_0 + \arg (z - z_1) + \arg (z - z_2) + \cdots + \arg (z - z_n).$$

Обозначим через $\Delta \arg f(z)$ приращение аргумента $f(z)$, вычисленного в предположении, что z обходит контур C. Очевидно, что $\Delta \arg f(z)$ составляет 2π, взятое столько раз, сколько раз точка $f(z)$ обходит вокруг начала.

Очевидно, что

$$\Delta \arg f(z) = \Delta \arg a_0 + \Delta \arg (z - z_1) + \Delta \arg (z - z_2) + \ldots + \Delta \arg (z - z_n).$$

Ясно, что $\Delta \arg a_0 = 0$, ибо a_0 — постоянная. Далее, $z - z_1$ изображается вектором, исходящим из точки z_1 в точку z. Допустим, что z_1 находится внутри области. Геометрически очевидно (рис. 9), что вектор $z - z_1$ при обходе точкой z контура C совершит один полный оборот вокруг своего начала, так что $\Delta \arg (z - z_1) = 2\pi$. Допустим теперь, что точка z_2 находится вне области. В этом случае вектор «качнется» в ту и другую сторону и вернется в исходное положение, не совершив оборота вокруг начала, так что $\Delta \arg (z = z_2) = 0$. Так можно рассуждать относительно

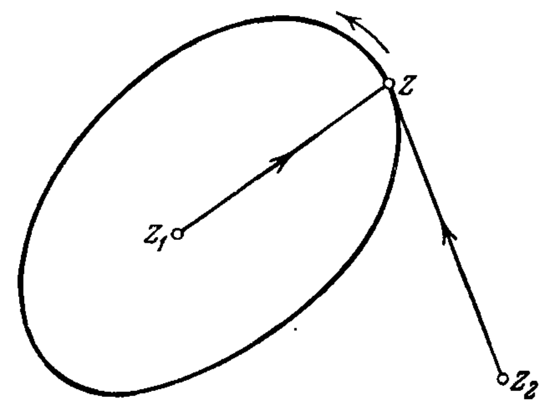


Рис. 9.

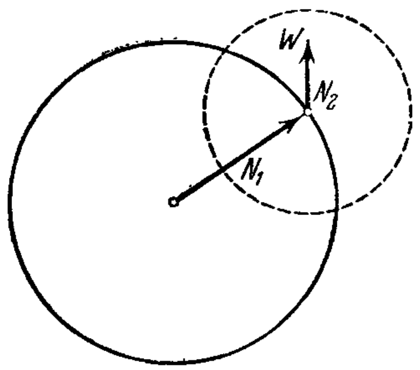


Рис. 10.

всех корней. Следовательно, $\Delta \arg f(z)$ равно 2π, умноженному на число корней $f(z)$, лежащих внутри области. Итак, число корней $f(z)$ внутри области равно числу обходов точки $f(z)$ вокруг начала, что и требовалось доказать.

Доказанная теорема дает возможность решить поставленную задачу в каждом частном случае, или нарисовать кривую, которую зачерчивает точка $f(z)$, можно с любой степенью точности. Для этого нужно взять достаточно густо совокупность точек z на контуре C, вычислить соответствующие значения $f(z)$ и соединить их непрерывной линией. Однако в некоторых случаях можно обойтись без этих утомительных вычислений. Укажем один из приемов на численном примере.

Пример. Узнать число корней многочлена $f(z) = z^{11} + 5z^2 - 2$ внутри окружности с радиусом, равным единице, и центром в начале координат.

На указанной окружности $|z| = 1$ из трех слагаемых, из которых составлен многочлен $f(z)$, одно, именно $5z^2$, преобладает на остальными. Действительно $|5z^2| = 5$, а $|z^{11} - 2| \leqslant |z|^{11} + 2 = 3$. Это обстоятельство дает возможность рассуждать так. Обозначим $z^{11} + 5z^2 - 2$ через w, $5z^2$ через N_1, $z^{11} - 2$ через N_2. Пока точка z обходит один раз единичную окружность, $N_1 = 5z^2$ обойдет окружность радиуса 5 два раза, ибо $|N_1| = 5$, $\arg N_1 = 2 \arg z$. Точка w «привязана» к точке N_1 вектором, длина которого $|N_2| \leqslant 3$, т. е. расстояние от точки w до точки N_1 все время меньше расстояния от N_1 до начала координат.

Следовательно, точка w, как бы она ни «вертелась» вокруг N_1 (рис. 10), не может «самостоятельно» обойти вокруг начала и обойдет вокруг начала столько раз, сколько точка N_1, т. е. два раза. Следовательно, число корней $f(z)$ внутри рассматриваемой области равно двум.

Задача Гурвица. В механике, именно в теории колебаний и регулирования, очень важны условия, позволяющие узнать, будут ли все корни данного многочлена $f(z) = a_0z^n + a_1z^{n-1} + \ldots + a_n$ (с действитель-

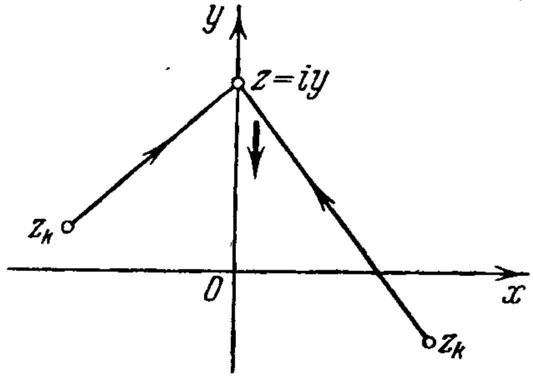


Рис. 11.

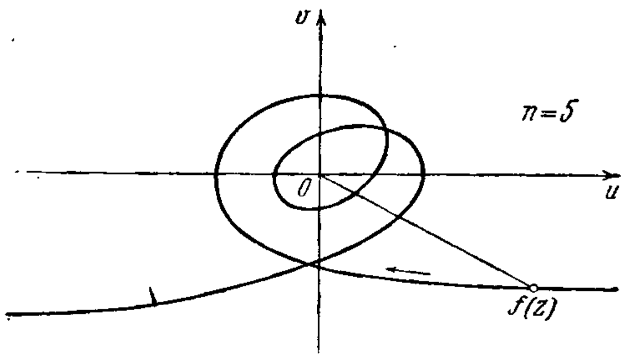


Рис. 12.

ными коэффициентами) иметь отрицательные действительные части, т. е. будут ли все корни расположены в полуплоскости налево от мнимой оси.

Один критерий для решения этой задачи легко получить из соображений, сходных с принципом аргумента.

Будем предполагать, что $a_0 > 0$.

Пусть точка z (рис. 11) проходит мнимую ось сверху вниз, т. е. пусть $z = iy$, причем y изменяется от $+\infty$ до $-\infty$, оставаясь действительным. Тогда $f(z)$ зачертит некоторую кривую линию, имеющую бесконечные ветви. Для исследования технически удобнее тесно связанная с ней кривая, зачерчиваемая функцией

$$f_1(z) = (i)^{-n}f(z) = a_0y^n - a_2y^{n-2} + a_4y^{n-4} + \ldots - i(a_1y^{n-1} - a_3y^{n-3} + \ldots) =$$
$$= \varphi(y) - i\psi(y),$$

где

$$\varphi(y) = a_0y^n - a_2y^{n-2} + \ldots,$$
$$\psi(y) = a_1y^{n-1} - a_3y^{n-3} + \ldots$$

Так как $\arg i = \frac{\pi}{2}$, то $\arg f_1(z) = -\frac{n\pi}{2} + \arg f(z)$ и, следовательно, приращения аргументов $f(z)$ и $f_1(z)$ одинаковы.

Подсчитаем приращение аргумента точки $f_1(z)$, пока z проходит сверху вниз мнимую ось.

Пусть $f(z) = a_0(z - z_1)(z - z_2)\ldots(z - z_n)$. Тогда

$$\arg f_1(z) = \arg(a_0i^{-n}) + \arg(z - z_1) + \arg(z - z_2) + \ldots + \arg(z - z_n).$$

Геометрически очевидно, что приращение $\arg(z - z_k)$ равно π, если z_k лежит в правой полуплоскости, и равно $-\pi$, если z лежит в левой полуплоскости (рис. 11).

Поэтому приращение аргумента $f_1(z)$ равно $\pi(N_1-N_2)$, где N_1 — число корней $f(z)$ в правой полуплоскости, N_2 — число корней в левой полуплоскости. Для того чтобы все корни лежали в левой полуплоскости, необходимо и достаточно, чтобы приращение аргумента точки $f_1(z)$ равнялось $-\pi n$, т. е. чтобы точка $f_1(z)$ совершила n полуоборотов по часовой стрелке вокруг начала координат (рис. 12).

Заметим, что точка $f_1(z)=\varphi(y)-i\psi(y)$ пересекает мнимую ось при значениях y, являющихся корнями $\varphi(y)$, а действительную ось — при корнях $\psi(y)$. Так как $\varphi(y)$ имеет не более n действительных корней, а число действительных корней $\psi(y)$ не более $n-1$, легко убедиться геометрически, что $f_1(z)$ может совершить n полных полуоборотов по часовой стрелке в том и только в том случае, если кривая идет из четвертого квадранта и затем пересекает по очереди отрицательную часть мнимой оси, отрицательную часть действительной оси, положительную часть мнимой оси, положительную часть действительной оси и т. д. так, что общее число точек пересечения с мнимой осью равно n (по одной на каждый полуоборот), а с действительной осью равно $n-1$ (на единицу меньше полуоборотов). Поэтому коэффициент a_1 должен быть положительным, а корни многочленов $\varphi(y)$ и $\psi(y)$ должны быть все действительными и все перемежающимися. Последнее означает, что если $y_1>y_2>\ldots>y_n$ — корни $\varphi(y)$, расположенные в порядке убывания, а $\eta_1>\eta_2>\ldots>\eta_{n-1}$ — корни $\psi(y)$, то $y_1>\eta_1>y_2>\eta_2\ldots>y_{n-1}>\eta_{n-1}>y_n$.

Итак, для того чтобы все корни многочлена $f(z)=a_0z^n+a_1z^{n-1}+\ldots+a_n$ с действительными коэффициентами и с $a_0>0$ лежали в левой полуплоскости, необходимо и достаточно, чтобы коэффициент a_1 был положительным и корни многочленов $\varphi(y)=a_0y^n-a_2y^{n-2}+a_4y^{n-4}-\ldots$ и $\psi(y)=a_1y^{n-1}-a_3y^{n-3}+\ldots$ были все действительными и перемежались.

Это условие равносильно известному условию Гурвица, заключающемуся в положительности определителей

$$a_1,\quad \begin{vmatrix} a_1 & a_0 \\ a_3 & a_2 \end{vmatrix},\quad \begin{vmatrix} a_1 & a_0 & a_{-1} \\ a_3 & a_2 & a_1 \\ a_5 & a_4 & a_3 \end{vmatrix},\quad \ldots,\quad \begin{vmatrix} a_1 & a_0 & a_{-1} & \ldots & a_{2-n} \\ a_3 & a_2 & a_1 & \ldots & a_{4-n} \\ \cdot & \cdot & \cdot & \ldots & \cdot \\ \cdot & \cdot & \cdot & \ldots & \cdot \\ \cdot & \cdot & \cdot & \ldots & \cdot \\ a_{2n-1} & a_{2n-2} & a_{2n-3} & \ldots & a_n \end{vmatrix}$$

в которых все a_i с номерами, меньшими нуля и бо́льшими n, должны быть заменены нулями (что такое определители, см. в главе XVI, том 3, § 3).

§ 5. ПРИБЛИЖЕННОЕ ВЫЧИСЛЕНИЕ КОРНЕЙ

Способ Штурма в соединении с использованием границы, меньше которой не может быть разность двух различных действительных корней, позволяет произвести «отделение» действительных корней многочлена с действительными коэффициентами, т. е. позволяет для каждого такого корня указать границы a и b, между которыми находится только один этот корень. Остается указать удобный способ для нахождения на отрезке $a < b$ чисел $\alpha_1 < \alpha_2 < \alpha_3 < \ldots$ и $\beta_1 > \beta_2 > \beta_3 > \ldots$, которые возможно быстрее подходили бы все ближе и ближе к величине искомого корня, причем первые являлись бы его приближениями по недостатку, а вторые — по избытку. Каждое из двух приближений α_k и β_k отличается от искомого корня x, очевидно, меньше, чем на их разность $\beta_k - \alpha_k$, так как корень лежит между ними. Таким образом, можно оценить, меньше чего заведомо будет погрешность, если остановиться на данном приближении.

График многочлена. Пусть задан многочлен n-й степени с действительными коэффициентами

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

Рассмотрим линию, выражаемую в прямоугольных координатах уравнением $y = f(x)$, т. е. график этого многочлена. Линию эту иногда называют параболой n-го порядка. Во-первых, очевидно, что для любого действительного x имеется одно и только одно определенное действительное $y = f(x)$, поэтому график f простирается сколь угодно далеко и направо и налево. Кроме того, при непрерывном изменении x как $f(x)$, так и $f'(x)$ изменяются непрерывно, т. е. без скачков. Поэтому график f есть плавная

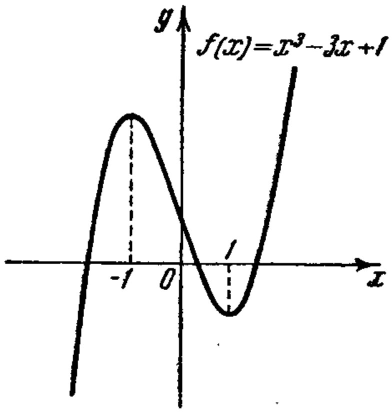


Рис. 13.

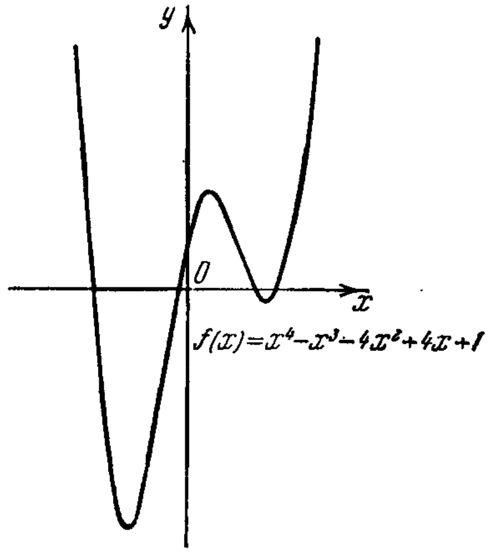


Рис. 14.

линия. При больших по абсолютной величине x первый член a_0x^n превышает по своей абсолютной величине сумму всех остальных членов, так как они все более низких степеней. Отсюда следует, что если n — четное и $a_0 > 0$, то график f и вправо и влево уходит на бесконечность кверху (а если $a_0 < 0$, то книзу); если же n — нечетное и $a_0 > 0$, то оно уходит вправо кверху, а влево книзу (если $a_0 < 0$, то наоборот).

Точки пересечения графика f с осью Ox, т. е. те точки, где $y = f(x) = 0$, соответствуют действительным корням уравнения $f(x) = 0$; их не больше, чем n. В максимумах и минимумах графика $y = f(x)$ производная $f'(x) = 0$, следовательно, число всех максимумов и минимумов не больше, чем $n - 1$. Если на некотором участке $f''(x) > 0$, то на нем первая производная возрастает, т. е. график обращен своею вогнутостью кверху; если $f''(x) < 0$, то он обращен вогнутостью книзу. Хотя бы уже потому, что некоторые корни $f'(x) = 0$ могут быть комплексными, число максимумов и минимумов графика f может быть и меньше, чем $n-1$.

Вот примеры графиков некоторых многочленов

$$f(x) = x^3 - 3x + 1 \quad (\text{рис. } 13),$$
$$f(x) = x^4 - x^3 - 4x^2 + 4x + 1 \quad (\text{рис. } 14).$$

Построив график многочлена, легко найти приближенно его корни. Именно, корни являются абсциссами точек пересечения графика с осью Ox.

Способ «недолета» и «перелета». Подставим в многочлен $f(x)$ какое-нибудь целое рациональное число, например 3, и будем затем подставлять 4, 5, ... Если при подстановке чисел 4, 5, 6 в результате получается еще тот же знак, как при подстановке числа 3, а при подстановке 7 — обратный, то ясно, что $f(x)$ имеет между 6 и 7 хотя бы один корень. Теперь будем подставлять 6; 6,1; 6,2;.... Так найдем два соседних из этой последовательности числа, например 6,4 и 6,5, которые при подстановке дают разные знаки. Значит, между ними имеется хотя бы один корень. Далее, подставляем 6,4; 6,41; 6,42; 6,43;... и находим еще более тесные границы для корня, например 6,42 и 6,43 и т. д. Это — способ «недолета - перелета». Этот способ можно еще значительно упростить, применяя на каждом шаге вычисления дополнительное преобразование многочлена, и тогда на каждом шаге, кроме первого, придется только подставлять целые числа, а не дроби, и притом только целые числа 1, 2, 3,...,9. Но мы не будем останавливаться на этом упрощении.

Способ касательных и способ хорд. Способ касательных, называемый способом Ньютона, и способ хорд, или прямолинейного интерполирования, называемый также способом ложного положения (regula falsi), используются либо раздельно, либо вместе для получения оценки погрешности. Пусть $a < b$ и между a и b имеется только один действительный корень многочлена $f(x)$ [знаки $f(a)$ и $f(b)$ различны], и, кроме того, вторая производная $f''(x)$ везде между a и b одного и того же знака. В таком случае часть графика f между $x = a$ и $x = b$ имеет один из четырех видов (рис. 15).

В случаях I и II касательная к графику в точке с абсциссой a пересекает ось Ox в точке с абсциссой α_1, лежащей между искомым корнем и a. Найдя абсциссу α_1 и рассмотрев далее касательную к графику в точке его с абсциссой α_1, мы аналогично найдем точку α_2, лежащую между точкой α_1 и искомым корнем; далее аналогично найдем точку α_3 и т. д. Так будут получаться всё лучшие приближения с недостатком. Как видно из чертежа, значения эти очень быстро приближаются к искомому корню.

В случаях *III* и *IV* надо, наоборот, начинать с абсциссы b, тогда будут получаться точки β_1, β_2, $\beta_3, \ldots$, т. е. всё лучшие приближения по избытку. Какой из четырех случаев имеет место, легко определить по знакам $f(a)$, $f(b)$ и $f''(x)$ для $a<x<b$.

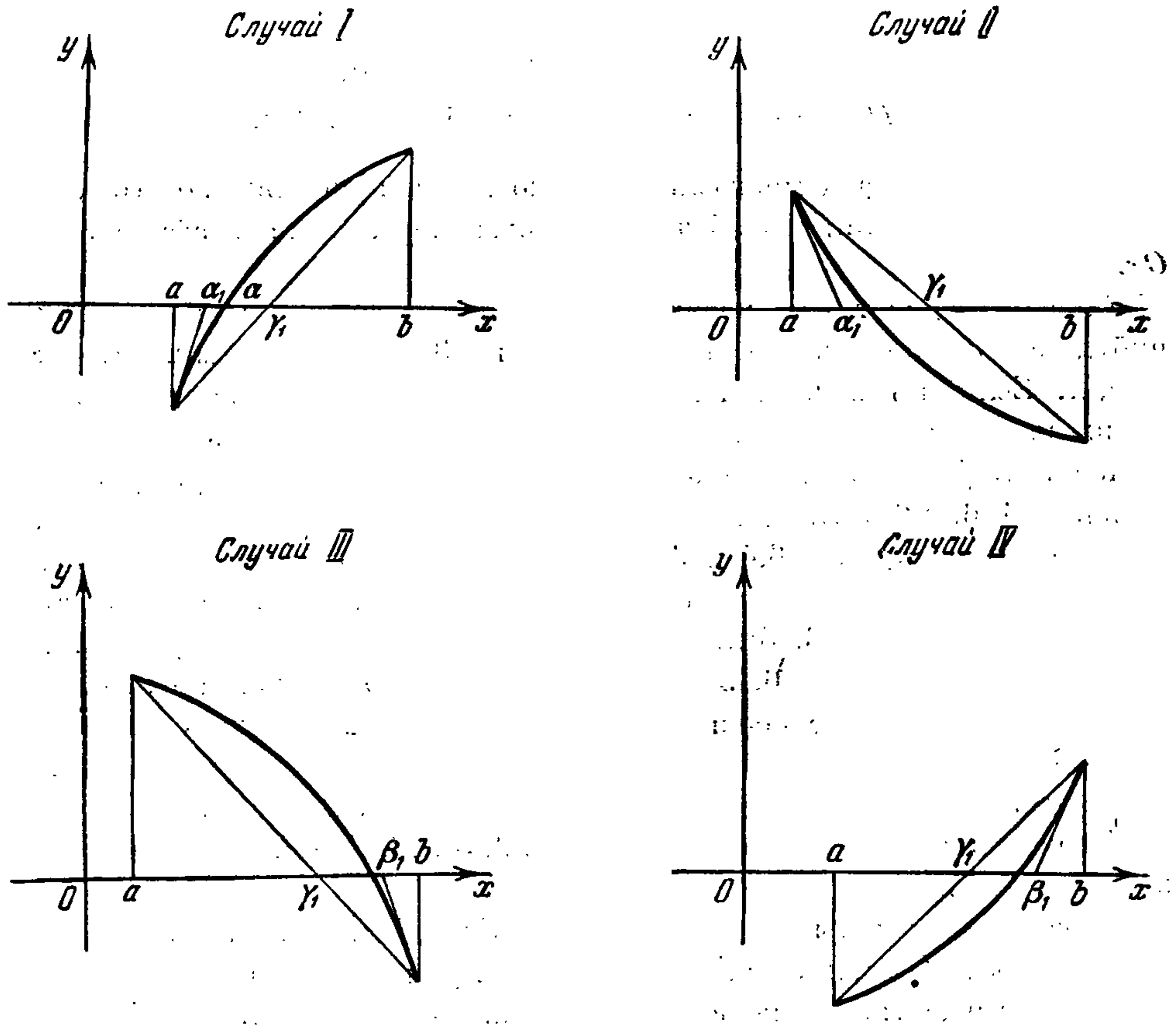


Рис. 15.

Так как уравнение касательной к линии $y = f(x)$ в точке ее с абсциссой a есть

$$y - f(a) = f'(a)(x - a),$$

то абсцисса α_1 точки ее пересечения с осью Ox получается из равенства

$$0 - f(a) = f'(a)(\alpha_1 - a)$$

равной

$$\alpha_1 = a - \frac{f(a)}{f'(a)}.$$

Далее,

$$\alpha_2 = \alpha_1 - \frac{f(\alpha_1)}{f'(\alpha_1)}, \quad \alpha_3 = \alpha_2 - \frac{f(\alpha_2)}{f'(\alpha_2)}$$

и т. д.

Аналогично

$$\beta_1 = b - \frac{f(b)}{f'(b)}, \quad \beta_2 = \beta_1 - \frac{f(\beta_1)}{f'(\beta_1)}, \quad \beta_3 = \beta_2 - \frac{f(\beta_2)}{f'(\beta_2)}$$

и т. д.

Это — способ Ньютона[1].

Способ прямолинейного интерполирования, или ложного положения, состоит в следующем. Уравнение хорды, как уравнение прямой, проходящей через две заданные точки, имеет вид

$$\frac{x-a}{b-a} = \frac{y-f(a)}{f(b)-f(a)},$$

и абсцисса γ_1 точки ее пересечения с осью Ox получается из уравнения

$$\frac{x-a}{b-a} = \frac{0-f(a)}{f(b)-f(a)}$$

равной

$$\gamma_1 = -\frac{(b-a)f(a)}{f(b)-f(a)} + a = \frac{af(b)-bf(a)}{f(b)-f(a)}.$$

Приняв ее за новое b в случаях I и II и за новое a в случаях III и IV, мы находим в случаях I и II

$$\gamma_2 = \frac{af(\gamma_1)-\gamma_1 f(a)}{f(\gamma_1)-f(a)}, \quad \gamma_3 = \frac{af(\gamma_2)-\gamma_2 f(a)}{f(\gamma_2)-f(a)}$$

и т. д.

В случаях III и IV принимаем γ_1 за новое a и получаем

$$\gamma_2 = \frac{\gamma_1 f(b)-bf(\gamma_1)}{f(b)-f(\gamma_1)}, \quad \gamma_3 = \frac{\gamma_2 f(b)-bf(\gamma_2)}{f(b)-f(\gamma_2)}$$

и т. д.

Особенно важно объединение обоих этих способов, ибо (как это видно на чертежах) оно позволяет, если известны приближения сверху и снизу, оценивать погрешность, которая, очевидно, не больше их разности, так как вычисляемый корень находится между ними.

З а м е ч а н и е. Важно заметить, что то обстоятельство, что $f(x)$ — именно многочлен, а не какая-либо другая функция от x, ни в способе Ньютона, ни в способе прямолинейного интерполирования не играет

[1] Из этих формул получаем также строгое доказательство обоих утверждений, сделанных выше из рассмотрения чертежа. Именно, значения α_n (аналогично и β_n) при увеличивающихся n изменяются монотонно, например в случае I увеличиваются и ограничены в своей совокупности, т. е., в силу леммы Вейерштрасса, приближаются к некоторому пределу α. Заменив же в этих формулах α_n его пределом α, получаем: $\alpha = \alpha - \frac{f(\alpha)}{f'(\alpha)}$, откуда $f(\alpha) = 0$, т. е. α — корень многочлена f.

никакой роли, т. е. оба эти способа и их объединение могут при вышеуказанных условиях применяться и для трансцендентных уравнений.

Способ Лобачевского. Одним из самых распространенных, в настоящее время способов вычислений корней, в особенности комплексных корней, является способ[1], предложенный Н. И. Лобачевским в его книге «Алгебра», вышедшей в 1834 г. Основная идея этого способа восходит еще к Бернулли.

Заметим, во-первых, что если дан многочлен, корни которого x_1, $x_2, \ldots, x_n$, то легко написать многочлен той же n-й степени, корнями которого будут x_1^2, $x_2^2, \ldots, x_n^2$, т. е. квадраты корней данного многочлена. Действительно, если x_1, $x_2, \ldots, x_n$ — корни многочлена

$$x^n + a_1x^{n-1} + a_2x^{n-2} + \ldots + a_n,$$

то он равен

$$(x - x_1)(x - x_2)\ldots(x - x_n);$$

многочлен же

$$x^n - a_1x^{n-1} + a_2x^{n-2} - \ldots \pm a_n,$$

корни которого — взятые с обратным знаком корни заданного многочлена, равен

$$(x + x_1)(x + x_2)\ldots(x + x_n).$$

Произведение этих двух многочленов есть, следовательно,

$$(x^2 - x_1^2)(x^2 - x_2^2)\ldots(x^2 - x_n^2)$$

и потому содержит только четные степени x. Если в нем положить $x^2 = y$, то мы получим многочлен n-й степени от y

$$y^n + b_1y^{n-1} + b_2y^{n-2} + \ldots + b_n,$$

равный

$$(y - x_1^2)(y - x_2^2)\ldots(y - x_n^2),$$

т. е. такой, корни которого суть x_1^2, $x_2^2, \ldots, x_n^2$. Вместо того, чтобы непосредственно умножать многочлен

$$x^n + a_1x^{n-1} + a_2x^{n-2} + \ldots + a_n$$

на многочлен

$$x^n - a_1x^{n-1} + a_2x^{n-2} - \ldots \pm a_n,$$

можно получить коэффициенты b_k по такой схеме. В первой строке над чертой пишутся 1, a_1, $a_2, \ldots, a_n$, а ниже, под чертой, под каждым из этих коэффициентов a_k — сначала его квадрат a_k^2, затем минус удвоенное

[1] Этот способ независимо друг от друга предложили Данделен (1826), Н. И. Лобачевский (1834) и Греффе (1837).

произведение его соседей

$$-2a_{k-1}a_{k+1},$$

затем плюс удвоенное произведение коэффициентов

$$+2a_{k-2}a_{k+2},$$

симметричных множителю a_k, и т. д., знакочередуясь, пока все дальнейшие коэффициенты с той или иной стороны не будут равны нулю. Тогда коэффициенты b_k получаются как суммы соответственных столбиков чисел, подписанных под чертой.

Получив так коэффициенты $1, b_1, b_2, \ldots, b_n$ многочлена, корнями которого будут $1, x_1^2, x_2^2, \ldots, x_n^2$, так же получают коэффициенты $1, c_1, c_2, \ldots, c_n$ многочлена, корнями которого будут квадраты корней многочлена

$$y^n + b_1 y^{n-1} + b_2 y^{n-2} + \ldots + b_n,$$

т. е. $x_1^4, x_2^4, \ldots, x_n^4$. Далее, аналогично можно получить коэффициенты $1, d_1, d_2, \ldots, d_n$ многочлена, корнями которого будут $x_1^8, x_2^8, \ldots, x_n^8$; затем многочлена с корнями $x_1^{16}, x_2^{16}, \ldots, x_n^{16}$ и т. д.

Рассмотрим только основную идею способа Лобачевского, причем ограничимся, для простоты, тем случаем, когда все корни уравнения действительны и различны по абсолютной величине. Пусть

$$|x_1| > |x_2| > \ldots > |x_n|,$$

т. е. x_1 — самый большой по абсолютной величине корень, x_2 — следующий и т. д. Пусть N — достаточно велико и многочлен

$$X^n + A_1 X^{n-1} + A_2 X^{n-2} + \ldots + A_n$$

имеет своими корнями N-ые степени корней $x_1, x_2, \ldots, x_n$ заданного многочлена, т. е.

$$\begin{aligned} -A_1 &= x_1^N + x_2^N + \ldots + x_n^N, \\ A_2 &= x_1^N x_2^N + x_1^N x_3^N + \ldots + x_{n-1}^N x_n^N, \\ & \ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ \pm A_n &= x_1^N x_2^N \ldots x_n^N. \end{aligned}$$

Тогда в ряду чисел $|x_1^N|, |x_2^N|, \ldots, |x_n^N|$, при больших N, каждое последующее число настолько меньше предшествующего, что в рассматриваемых выражениях для $A_1, A_2, \ldots, A_n$ можно оставить только первое слагаемое, а суммой всех остальных слагаемых по отношению к нему пренебречь. Тогда мы получим приближенные формулы

$$x_1^N \approx -A_1, \qquad x_1^N x_2^N \approx A_2,$$

$$x_1^N x_2^N x_3^N \approx A_3, \quad \ldots, \quad x_1^N x_2^N x_3^N \ldots x_n^N \approx \pm A_n,$$

или, деля их попарно друг на друга и извлекая корень N-й степени, такие формулы для самих x_k:

$$x_1 = \sqrt[N]{-A_1}, \quad x_2 = \sqrt[N]{-\frac{A_2}{A_1}}, \quad x_3 = \sqrt[N]{-\frac{A_3}{A_2}}, \ldots, x_n = \sqrt[N]{-\frac{A_n}{A_{n-1}}}.$$

Можно показать, что продолжать вычисление достаточно до того многочлена, коэффициенты которого, взятые со знаками $+ - + - \ldots$, будут, с нужной степенью точности, квадратами соответственных коэффициентов предыдущего многочлена.

Подробное изложение способа Лобачевского можно найти в известной книге академика А. Н. Крылова «Лекции о приближенных вычислениях».

ЛИТЕРАТУРА

Популярные брошюры

Курош А. Г. Алгебраические уравнения произвольных степеней. Гостехиздат, 1951.

Маркушевич А. И. Комплексные числа и конформные отображения. Гостехиздат, 1954.

Шафаревич И. Р. О решении уравнений высших степеней (Метод Штурма), Гостехиздат, 1954.

Систематические университетские курсы

Курош А. Г. Курс высшей алгебры. Гостехиздат, 1952.

Окунев Л. Я. Высшая алгебра. Гостехиздат, 1949. (Курс для студентов пединститутов).

Сушкевич А. К. Основы высшей алгебры. ГОНТИ, 1937.

В первой из этих книг более подчеркнута абстрактная сторона алгебры, а в двух других уделяется большее внимание конкретному содержанию.

ИМЕННОЙ УКАЗАТЕЛЬ

ОГЛАВЛЕНИЕ

СОДЕРЖАНИЕ ВТОРОГО И ТРЕТЬЕГО ТОМОВ

ТОМ II

ТОМ III

Математика, ее содержание, методы и значение
Том I
*
Утверждено к печати Математическим институтом им. В. А. Стеклова Академии наук СССР

Редактор издательства *А. З. Рывкин*. Технический редактор *Е. Б. Поливанова*

РИСО АН СССР № 28-13В. Сдано в набор 1/III 1956 г. Подп. в печать 18/VII 1956 г.
Формат бумаги $70\times108^1/_{16}$. Печ. л. 18,5—25,4. Уч.-изд. лист. 20,5. Тираж 7000 экз. Т-06757.
Изд. № 1590. Тип. зак. 210 *Цена 15 р. 85 к.*

Издательство Академии наук СССР. Москва, Б-64, Подсосенский пер., д. 21

2-я типография Издательства АН СССР. Москва, Г-99, Шубинский пер., д. 10

ОПЕЧАТКИ

Страница	Строка	Напечатано	Должно быть
60	5 св.	самых	самих
73	16 св.	различие	развитие
88	8-10 св.	$\overline{\overline{s}}$	h
181	8 сн.	уравнениям	уравнения
194	18 сн.	$r_n^2 m_u$	$r_n^2 m_n$

Математика, т. I