К А Д Е М И Я Н А У К С С С
МАТЕМАТИЧЕСКИЙ ИНСТИТУТ им. В. А. СТЕКЛОВА

МАТЕМАТИКА,
ЕЕ СОДЕРЖАНИЕ, МЕТОДЫ И ЗНАЧЕНИЕ

ТОМ ТРЕТИЙ

ИЗДАТЕЛЬСТВО АКАДЕМИИ НАУК СССР
МОСКВА 1956

Глава XV

ТЕОРИЯ ФУНКЦИЙ ДЕЙСТВИТЕЛЬНОГО ПЕРЕМЕННОГО

§ 1. ВВЕДЕНИЕ

К концу XVIII — началу XIX в. дифференциальное и интегральное исчисление было в основном разработано. До этого времени (фактически, весь XVIII век) ученые были заняты построением его отдельных разделов, открывали все новые и новые факты, развивали все новые и новые области приложений дифференциального и интегрального исчисления к различным вопросам механики, астрономии, техники. Теперь появилась возможность обозреть полученные результаты, заняться их систематизацией, вникнуть в смысл основных понятий анализа. И вот выясняется, что с основами анализа дело обстоит не совсем благополучно.

Еще в XVIII в. у крупнейших математиков того времени не было единого мнения насчет того, что такое функция. Это приводило к долгим спорам о том, правильно или неправильно то или иное решение задачи, правилен или неправилен тот или иной конкретный математический результат. Постепенно выяснилось, что и другие основные понятия анализа нуждаются в уточнении. Недостаточно четкое понимание того, что такое непрерывность и каковы свойства непрерывных функций, привело к появлению ряда ошибочных утверждений, например, что непрерывная функция всегда дифференцируема. Математика стала оперировать со столь сложными функциями, что стало уже невозможно ссылаться на очевидность и догадку. Появилась настоятельная необходимость навести порядок в основных понятиях анализа.

Первая серьезная попытка в этом направлении была предпринята Лагранжем, а затем на тот же путь встал Коши. Коши уточнил и ввел во всеобщее употребление сохранившиеся до наших дней определения предела, непрерывности, интеграла. Примерно в то же время чешский математик Больцано провел строгое изучение основных свойств непрерывных функций.

Рассмотрим эти свойства непрерывных функций более подробно. Пусть непрерывная функция $f(x)$ задана на некотором отрезке $[a, b]$, т. е. для всех чисел, удовлетворяющих неравенствам $a \leqslant x \leqslant b$. Ранее считалось очевидным, что если на концах отрезка функция принимает

значения разных знаков, то в некоторой промежуточной точке она обращается в нуль. Теперь этот факт получил строгое обоснование. Точно так же было строго доказано, что непрерывная функция, заданная на отрезке, принимает в некоторых точках свое наибольшее и наименьшее значение.

Исследование этих свойств непрерывных функций заставило глубже вникнуть в природу действительных чисел. В результате появилась теория действительных чисел, были четко сформулированы основные свойства числовой прямой.

Дальнейшее развитие математического анализа привело к необходимости рассматривать все более и более «плохие», в частности разрывные, функции. Разрывные функции появляются, например, как пределы непрерывных функций, причем заранее не известно, будет ли предельная функция непрерывной, или нет, а также при схематизации процессов с резким внезапным изменением. Возникла новая задача — обобщение аппарата анализа на разрывные функции.

Риман исследовал вопрос о том, на какие классы разрывных функций распространяется понятие интеграла. В результате всей этой деятельности по обоснованию анализа появилась новая математическая дисциплина: теория функций действительного переменного.

Если классический математический анализ оперирует в основном с «хорошими» (например, непрерывными, дифференцируемыми) функциями, то теория функций действительного переменного изучает значительно более общие классы функций. Если в математическом анализе дается определение какой-либо операции (например, интегрирования) для непрерывных функций, то для теории функций действительного переменного характерно исследование вопроса о том, для какого класса функций применимо это определение, как следует видоизменить определение, чтобы оно стало более широким. В частности, лишь теория функций действительного переменного смогла дать удовлетворительный ответ на вопрос о том, что такое длина кривой и для каких кривых имеет смысл говорить о длине.

Основанием, на котором строится сама теория функций действительного переменного, является *теория множеств*.

В соответствии с этим мы начинаем наше изложение с рассмотрения элементов теории множеств, затем переходим к изучению точечных множеств и завершаем главу изложением одного из основных понятий теории функций действительного переменного, а именно интеграла Лебега.

§ 2. МНОЖЕСТВА

Людям постоянно приходится иметь дело с различными совокупностями предметов. Как уже разъяснялось в главе I (том 1), это повлекло за собой возникновение понятия *числа*, а затем и понятия

множества, которое является одним из основных простейших математических понятий и не поддается точному определению. Нижеследующие замечания имеют своей целью пояснить, что такое множество, но не претендуют на то, чтобы служить его определением.

Множеством называется собрание, совокупность, коллекция вещей, объединенных по какому-либо признаку или по какому-либо правилу. Понятие множества возникает путем абстракции. Рассматривая какую-либо совокупность предметов как множество, отвлекаются от всех связей и соотношений между различными предметами, составляющими множества, но сохраняют за предметами их индивидуальные черты. Таким образом, множество, состоящее из пяти монет, и множество, состоящее из пяти яблок, — это разные множества. С другой стороны, множество из пяти монет, расположенных по кругу, и множество из тех же монет, положенных одна на другую, — это одно и то же множество.

Приведем несколько примеров множеств. Можно говорить о множестве песчинок, составляющих кучу песка, о множестве всех планет нашей солнечной системы, о множестве всех людей, находящихся в данный момент в каком-либо доме, о множестве всех страниц этой книги. В математике тоже постоянно встречаются различные множества, например множество всех корней заданного уравнения, множество всех натуральных чисел, множество всех точек на прямой и т. д.

Математическая дисциплина, изучающая общие свойства множеств, т. е. свойства множеств, не зависящие от природы составляющих их предметов, называется *теорией множеств*. Эта дисциплина начала бурно развиваться в конце XIX и начале XX в. Основатель научной теории множеств — немецкий математик Г. Кантор.

Работы Кантора по теории множеств выросли из рассмотрения вопросов сходимости тригонометрических рядов. Это весьма обычное явление: очень часто рассмотрение конкретных математических задач ведет к построению весьма абстрактных и общих теорий. Значение таких абстрактных построений определяется тем, что они оказываются связанными не только с той конкретной задачей, из которой они выросли, но имеют приложения и в ряде других вопросов. В частности, именно так обстоит дело и с теорией множеств. Идеи и понятия теории множеств проникли буквально во все разделы математики и существенно изменили ее лицо. Поэтому нельзя получить правильного представления о современной математике, не познакомившись с элементами теории множеств. Особенно большое значение имеет теория множеств для теории функций действительного переменного.

Множество считается заданным, если относительно любого предмета можно сказать, принадлежит он множеству или не принадлежит. Иными словами, множество вполне определяется заданием всех

принадлежащих ему предметов. Если множество M состоит из предметов $a, b, c, \ldots$, и только из этих предметов, то пишут

$$M=\{a, b, c, \ldots\}.$$

Предметы, составляющие какое-либо множество, принято называть его *элементами*. Тот факт, что предмет m является элементом множества M, записывается в виде

$$m \in M$$

и читается: «m принадлежит M», или «m есть элемент M». Если же предмет n не принадлежит множеству M, то пишут: $n \overline{\in} M$. Каждый предмет может служить лишь одним элементом заданного множества; иными словами, все элементы одного и того же множества отличны друг от друга.

Элементы множества M могут сами быть множествами, однако, во избежание противоречий, приходится требовать, чтобы само множество M не было одним из своих собственных элементов: $M \overline{\in} M$.

Множество, не содержащее ни одного элемента, называется *пустым множеством*. Например, множество всех действительных корней уравнения

$$x^2+1=0$$

есть пустое множество. Пустое множество в дальнейшем будем обозначать через 0.

Если для двух множеств M и N каждый элемент x множества M является также элементом множества N, то говорят, что M входит в N, что M есть часть N, что M есть подмножество N или что M содержится в N; это записывается в виде

$$M \subseteq N \quad \text{или} \quad N \supseteq M.$$

Например, множество $M=\{1, 2\}$ есть часть множества $N=\{1, 2, 3\}$.

Ясно, что всегда $M \subseteq M$. Удобно считать, что пустое множество есть часть любого множества.

Два множества *равны*, если они состоят из одних и тех же элементов. Например, множество корней уравнения $x^2-3x+2=0$ и множество $M=\{1, 2\}$ между собою равны.

Определим правила *действий* над множествами.

Объединение или сумма. Пусть имеются множества $M, N, P, \ldots$ Объединением или суммой этих множеств называется множество X, состоящее из всех элементов, принадлежащих хотя бы одному из «слагаемых» $M, N, P, \ldots$

$$X=M+N+P+\ldots$$

При этом, даже если элемент x принадлежит нескольким слагаемым, то он входит в сумму X лишь один раз. Ясно, что

$$M+M=M,$$

и если $M \subseteq N$, то

$$M+N=N.$$

Пересечение. Пересечением или общей частью множеств M, N, $P, \ldots$ называется множество Y, состоящее из всех тех элементов, которые принадлежат одновременно всем множествам M, N, $P, \ldots$,

Ясно, что $M \cdot M = M$, и если $M \subseteq N$, то $M \cdot N = M$.

Если пересечение множеств M и N пусто: $M \cdot N = 0$, то говорят, что эти множества *не пересекаются*.

Для обозначения операции суммы и пересечения множеств употребляют также знаки $\sum$ и $\prod$. Таким образом,

$$E = \sum E_i$$

есть сумма множеств E_i, а

$$F = \prod E_i$$

— их пересечение.

Читателю рекомендуется доказать, что сумма и пересечение множеств связаны обычным распределительным законом

$$M(N+P) = MN + MP,$$

а также законом

$$M + NP = (M+N)(M+P).$$

Разность. Разностью двух множеств M и N называется множество Z всех тех элементов из M, которые не принадлежат N:

$$Z = M - N.$$

Если $N \subseteq M$, то разность $Z = M - N$ называют также *дополнением* к множеству N относительно M.

Нетрудно показать, что всегда

$$M(N-P) = MN - MP$$

и

$$(M-N) + MN = M.$$

Таким образом, правила действий над множествами значительно отличаются от обычных правил арифметики.

Конечные и бесконечные множества. Множества, состоящие из конечного числа элементов, называются конечными множествами. Если же число элементов множества неограниченно, то такое множество

называется бесконечным. Например, множество всех натуральных чисел бесконечно.

Рассмотрим два каких-либо множества M и N и поставим вопрос о том, одинаково или нет количество элементов в этих множествах.

Если множество M конечно, то количество его элементов характеризуется некоторым натуральным числом — числом его элементов. В этом случае для сравнения количества элементов множеств M и N достаточно сосчитать число элементов в M, число элементов в N и сравнить полученные числа. Естественно также считать, что если одно из множеств M и N конечно, а другое бесконечно, то бесконечное множество содержит больше элементов, чем конечное.

Однако, если оба множества M и N бесконечны, то путь простого счета элементов ничего не дает. Поэтому сразу возникают такие вопросы: все ли бесконечные множества имеют одинаковое количество элементов, или же существуют бесконечные множества с бо́льшим и меньшим количеством элементов? Если верно второе, то каким способом можно сравнивать между собой количество элементов в бесконечных множествах? Этими вопросами мы теперь и займемся.

Взаимно однозначное соответствие. Пусть снова M и N — два конечных множества. Как узнать, какое из этих множеств содержит больше элементов, не считая числа элементов в каждом множестве? Для этого будем составлять *пары*, объединяя в пару один элемент из M и один элемент из N. Тогда, если какому-нибудь элементу из M не найдется парного к нему элемента из N, то в M больше элементов, чем в N. Поясним это рассуждение примером.

Пусть в зале находится некоторое число людей и некоторое число стульев. Чтобы узнать, чего больше, достаточно попросить людей занять места. Если кто-нибудь остался без места, значит, людей больше, а если, скажем, все сидят и заняты все места, то людей столько же, сколько стульев. Описанный способ сравнения количества элементов во множествах имеет то преимущество перед непосредственным счетом элементов, что он без особых изменений применяется не только к конечным, но и к бесконечным множествам.

Рассмотрим множество всех натуральных чисел

$$M = \{1,\ 2,\ 3,\ 4,\ \ldots\}$$

и множество всех четных чисел

$$N = \{2,\ 4,\ 6,\ 8,\ \ldots\}.$$

Какое множество содержит больше элементов? На первый взгляд кажется, что первое. Однако мы можем образовать из элементов этих множеств пары, как указано ниже.

Таблица 1

M	1	2	3	4	. . .
N	2	4	6	8	. . .

Ни один элемент M и ни один элемент N не остается без пары. Правда, мы могли бы также образовать пары и так:

Таблица 2

M	1	2	3	4	5	. . .
N	—	2	—	4	—	. . .

Тогда многие элементы из M остаются без пар. С другой стороны, мы могли бы составить пары и так:

Таблица 3

M	—	1	—	2	—	3	—	. . .
N	2	4	6	8	10	12	14	. . .

Теперь многие элементы из M остаются без пар.

Таким образом, если множества A и B бесконечны, то различным способам образования пар соответствуют разные результаты. Если существует такой способ образования пар, при котором у каждого элемента A и каждого элемента B имеется парный к нему элемент, то говорят, что между множествами A и B можно установить взаимно однозначное соответствие. Например, между рассмотренными выше множествами M и N можно установить взаимно однозначное соответствие, как это видно из табл. 1.

Если между множествами A и B можно установить взаимно однозначное соответствие, то говорят, что они имеют *одинаковое* количество элементов или *равномощны*. Если же при *любом* способе образования пар некоторые элементы из A всегда остаются без пар, то говорят, что множество A содержит больше элементов, чем B, или что множество A имеет бóльшую мощность, чем B.

Таким образом, мы получили ответ на один из поставленных выше вопросов: как сравнивать между собой количество элементов в бесконечных множествах. Однако это нисколько не приблизило нас к ответу на другой вопрос: существуют ли вообще бесконечные множества,

имеющие различные мощности? Чтобы получить ответ на этот вопрос, исследуем некоторые простейшие типы бесконечных множеств.

Счетные множества. Если можно установить взаимно однозначное соответствие между элементами множества A и элементами множества всех натуральных чисел

$$Z = \{1, 2, 3. \ldots\},$$

то говорят, что множество A *счетно*. Иными словами, множество A счетно, если все его элементы можно занумеровать посредством натуральных чисел, т. е. записать в виде последовательности

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

Табл. 1 показывает, что множество всех четных чисел счетно (верхнее число рассматривается теперь как номер соответствующего нижнего числа).

Счетные множества это, так сказать, самые маленькие из бесконечных множеств: во всяком бесконечном множестве содержится счетное подмножество.

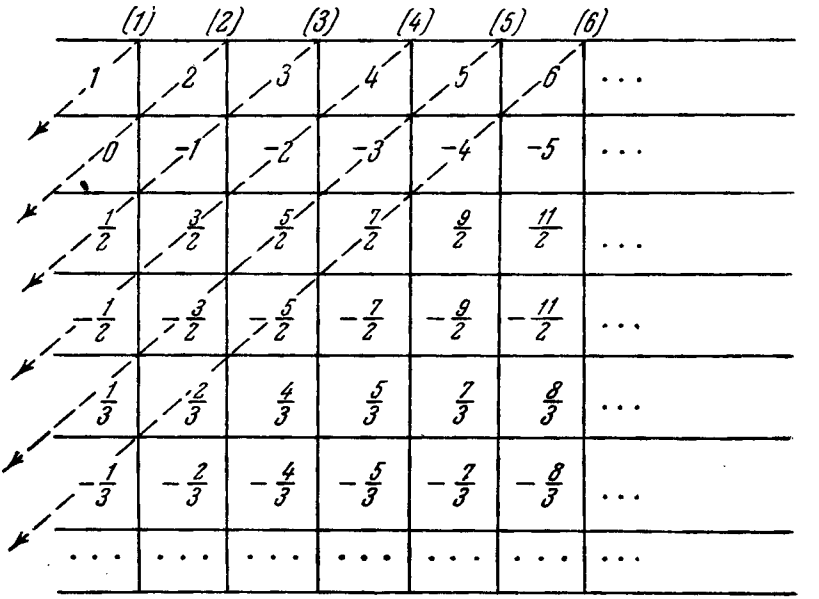
Если два непустых конечных множества не пересекаются, то их сумма содержит больше элементов, чем каждое из слагаемых. Для бесконечных множеств это правило может и не выполняться. В самом деле, пусть $Ч$ есть множество всех четных чисел, H — множество всех нечетных чисел и Z — множество всех натуральных чисел. Как показывает табл. 4, множества $Ч$ и H счетны. Однако множество $Z = Ч + H$ вновь счетно.

Таблица 4

$Ч$	2	4	6	8	. . .
H	1	3	5	7	. . .
Z	1	2	3	4	. . .

Нарушение правила «целое больше части» для бесконечных множеств показывает, что свойства бесконечных множеств *качественно* отличны от свойств конечных множеств. Переход от конечного к бесконечному сопровождается в полном согласии с известным положением диалектики — качественным изменением свойств.

Докажем, что *множество всех рациональных чисел счетно*. Для этого расположим все рациональные числа в такую таблицу:



Здесь в первой строке помещены все натуральные числа в порядке их возрастания, во второй строке 0 и целые отрицательные числа в порядке их убывания, в третьей строке — положительные несократимые дроби со знаменателем 2 в порядке их возрастания, в четвертой строке — отрицательные несократимые дроби со знаменателем 2 в порядке их убывания и т. д. Ясно, что каждое рациональное число один и только один раз находится в этой таблице. Перенумеруем теперь все числа этой таблицы в том порядке, как это указано стрелками. Тогда все рациональные числа разместятся в порядке одной последовательности:

Номер места, занимаемого рациональным числом	1	2	3	4	5	6	7	8	9 ...
Рациональное число	1,	2,	0,	3,	−1,	$\frac{1}{2}$,	4,	−2,	$\frac{3}{2}$, ...

Этим установлено взаимно однозначное соответствие между всеми рациональными числами и всеми натуральными числами. Поэтому множество всех рациональных чисел счетно.

Множества мощности континуума. Если можно установить взаимно однозначное соответствие между элементами множества M и точками отрезка $0 \leqslant x \leqslant 1$, то говорят, что множество M имеет *мощность континуума*. В частности, согласно этому определению, само множество точек отрезка $0 \leqslant x \leqslant 1$ имеет мощность континуума.

Из рис. 1 видно, что множество точек любого отрезка AB имеет мощность континуума. Здесь взаимно однозначное соответствие устанавливается геометрически, посредством проектирования.

Нетрудно показать, что множества точек любого интервала $a<x<b$ и всей числовой прямой $-\infty<x<+\infty$ имеют мощность континуума.

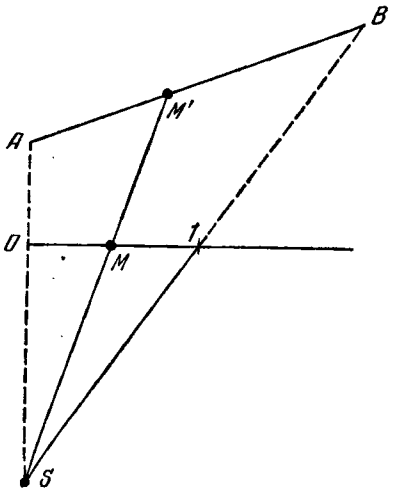


Рис. 1.

Значительно более интересен такой факт: множество точек квадрата $0\leqslant x\leqslant 1$, $0\leqslant y\leqslant 1$ имеет мощность континуума. Таким образом, грубо говоря, в квадрате «столько же» точек, сколько и в отрезке.

§ 3. ДЕЙСТВИТЕЛЬНЫЕ ЧИСЛА [1]

Развитие понятия о числе подробно изложено в главе I (том 1). Здесь мы дадим читателю беглое представление о теориях действительных чисел, которые возникли в XIX в. в связи с обоснованием основных понятий анализа.

Рациональные числа. Мы предполагаем, что читатель знаком с основными свойствами рациональных чисел. Не вдаваясь в подробности, напомним эти свойства. Рациональные числа, т. е. числа вида $\frac{m}{n}$, где m и n — целые и $n\neq 0$, представляют собою множество чисел, в котором определены две операции (сложение и умножение). Эти операции подчиняются ряду законов (аксиом). В дальнейшем a, b, c, ... обозначают рациональные числа.

I. Аксиомы сложения.

1) $a+b=b+a$ (коммутативность),

2) $a+(b+c)=(a+b)+c$ (ассоциативность),

3) уравнение

$$a+x=b$$

имеет единственное решение (существование обратной операции).

Из этих аксиом непосредственно вытекает, что имеет однозначный смысл выражение $a+b+c$, что существует рациональное число 0 (нуль), для которого $a+0=a$, и что для сложения существует обратная операция — вычитание, так что имеет смысл выражение $b-a$.

Таким образом, с алгебраической точки зрения, по отношению к операции сложения все рациональные числа образуют коммутативную группу.

---

[1] При написании этого параграфа использованы ценные консультации А. Н. Колмогорова.

II. Аксиомы умножения.

1) $ab = ba$ (коммутативность),

2) $a(bc) = (ab)c$ (ассоциативность),

3) уравнение

$$ay = b,$$

где $a \neq 0$, имеет единственное решение (существование обратной операции).

Из этих аксиом вытекает, что имеет смысл выражение abc, что существует рациональное число 1, для которого $a \cdot 1 = a$ и что для рациональных чисел, отличных от 0, существует обратная операция — деление. Все рациональные числа, исключая число 0, образуют коммутативную группу по отношению к операции умножения.

III. Аксиома дистрибутивности.

1) $(a+b)c = ac + bc$.

Все аксиомы I—III показывают, что по отношению к операциям сложения и умножения рациональные числа образуют так называемое *алгебраическое поле*.

IV. Аксиомы упорядоченности.

1) Для любых двух рациональных чисел a и b имеет место одно и только одно из трех соотношений: либо $a < b$, либо $a > b$, либо $a = b$.

2) Если $a < b$ и $b < c$, то $a < c$.

3) Если $a < b$, то $a + c < b + c$ (монотонность сложения).

4) Если $a < b$ и $c > 0$, то $ac < bc$ (монотонность умножения на $c > 0$).

Все эти аксиомы позволяют назвать множество рациональных чисел *упорядоченным полем*.

Кроме рациональных чисел, существуют и другие системы объектов, которые удовлетворяют этим аксиомам и, следовательно, являются упорядоченными полями.

Отметим два важных свойства рациональных чисел.

Плотность: для любых a и b, $a < b$, найдется такое c, что $a < c < b$.

Счетность: множество всех рациональных чисел счетно (см. § 2).

Об измерении величин. Недостаточность одних лишь рациональных чисел для математики проявляется уже при рассмотрении такой важной задачи, как задача об измерении величин. Мы рассмотрим ее на простейшем примере задачи об измерении длин отрезков.

Представим себе прямую, на которой отмечены определенное направление, начало отсчета (точка 0) и единица масштаба. Тогда понятно, что такое отрезок OA с концом в точке $\frac{1}{2}$, $\frac{1}{3}$, $\frac{2}{3}$, $-\frac{1}{3}$ и т. д. Вообще, каждому рациональному числу a можно поставить в соответствие точку A на прямой, а именно точку с координатой

$x=a$. В таком случае число a определяет длину направленного отрезка OA. Однако при такой конструкции длина не всякого отрезка измеряется некоторым (рациональным) числом. Например, как это было известно уже древним грекам, длина диагонали квадрата со стороной, равной единице, не измеряется никаким рациональным числом. Иначе говоря, точек на прямой больше, чем рациональных точек. Естественный выход из этого положения — установление взаимно однозначного соответствия между числами и длинами, т. е. дальнейшее расширение понятия числа.

Действительные числа. Мы пришли к выводу, что одних рациональных чисел для измерения величин недостаточно и что понятие числа должно быть расширено таким образом, чтобы между числами и точками на прямой существовало взаимно однозначное соответствие. С этой целью постараемся выяснить, нельзя ли определить положение произвольной точки на прямой при помощи одних лишь рациональных точек. Аналогичная конструкция в области рациональных чисел и приведет нас к понятию действительного числа.

Пусть α — произвольная точка на прямой. Тогда все рациональные точки a можно разделить на две части: к одной части отнесем все те точки a, которые расположены левее α, а к другой — те точки a, которые находятся правее α. Что же касается самой точки α (если она случайно оказалась рациональной), то ее можно отнести к любой из частей. Такое разбиение рациональных точек принято называть *сечением*. Сечения будем считать тождественными, если совокупности рациональных точек, входящих в левые и правые части сечений, совпадают (с точностью до одной точки). Теперь нетрудно видеть, что различные точки α и β определяют разные сечения. В самом деле, так как рациональные точки расположены на прямой всюду плотно, то найдутся рациональные точки r_1 и r_2, расположенные строго между α и β. Тогда для одного сечения они попадут в его правую часть, а для другого — в левую.

Итак, каждая точка на прямой определяет сечение в области рациональных точек и разным точкам соответствуют разные сечения. Очень важно, что сечения можно определить и несколько иначе, чем это было сделано выше, и притом так, чтобы само число α не фигурировало в этом определении. Именно, будем называть сечением в области рациональных точек такое разбиение всех рациональных точек на два непустых непересекающихся множества A и B, таких, что $a<b$ для любых $a\in A$, $b\in B$. При таком определении по сечению можно однозначно восстановить ту точку (рубеж), которая производит его. Иными словами, при помощи сечений в области рациональных точек можно определить *любую* точку на прямой. Изложенная конструкция была предложена немецким математиком Р. Дедекиндом и носит название *дедекиндова сечения*.

Сечения — не единственный возможный способ определения положения любой точки при помощи рациональных точек. К обычной практике измерений ближе лежит следующий способ Г. Кантора. Пусть снова α — произвольная точка на прямой. Тогда можно найти две сколь угодно близкие рациональные точки a и b, такие, что α заключено между a и b. Точки a и b определяют приближенно положение точки α. Представим себе этот процесс приближенного определения точки α неограниченно продолженным, и притом так, что на каждом следующем шаге его точность все более и более увеличивается. Тогда мы получим систему отрезков $[a_n, b_n]$ с концами в рациональных точках, таких, что $[a_{n+1}, b_{n+1}] \subseteq [a_n, b_n]$ и $b_n - a_n \to 0$ $(n \to \infty)$. Система отрезков, удовлетворяющая этим условиям, называется *стягивающейся системой отрезков*. Ясно, что такая система отрезков однозначно определяет положение точки α.

При помощи аналогичных построений в области рациональных чисел можно определить действительные числа. Далее определяются действия между действительными числами и устанавливается, что они удовлетворяют тем же аксиомам, что и действия над рациональными числами. Теперь уже каждой точке на прямой соответствует действительное число и обратно. В силу этого совокупность всех действительных чисел часто называют числовой прямой.

Принципы непрерывности. Между множеством всех рациональных чисел и множеством всех действительных чисел имеется существенное различие. Именно, совокупность всех действительных чисел обладает рядом свойств, характеризующих *непрерывность* этого множества, в то время как множество всех рациональных чисел такими свойствами не обладает. Эти свойства принято называть принципами непрерывности. Перечислим здесь важнейшие из них.

Принцип Дедекинда. Если множество всех действительных чисел разбито на два непустых множества X и Y без общих элементов, так что для любых $x \in X$, $y \in Y$ выполняется неравенство $x < y$, то существует единственное число ξ (рубеж), для которого $x \leqslant \xi \leqslant y$ для любых $x \in X$, $y \in Y$. Множество действительных чисел x, удовлетворяющих неравенствам $a \leqslant x \leqslant b$, называется отрезком числовой прямой и обозначается через $[a, b]$. Система отрезков $[a_n, b_n]$ называется *стягивающейся*, если $[a_{n+1}, b_{n+1}] \subseteq [a_n, b_n]$ и $b_n - a_n \to 0$ $(n \to \infty)$.

Принцип Кантора. Для любой стягивающейся системы отрезков $[a_n, b_n]$ существует одно и только одно действительное число ξ, принадлежащее каждому из этих отрезков.

Принцип Вейерштрасса. Всякая неубывающая ограниченная сверху последовательность действительных чисел сходится.

Будем говорить, что последовательность действительных чисел $\{x_n\}$ фундаментальна, если для любого $\varepsilon > 0$ найдется такое натуральное число N, что для всех $n > N$ и всех натуральных p

$$|x_{n+p} - x_n| < \varepsilon.$$

Принцип Коши. Всякая фундаментальная последовательность действительных чисел сходится.

Поскольку мы не провели аккуратного построения действительных чисел, мы не имеем возможности установить, что для множества действительных чисел выполняются эти принципы. Наша ближайшая цель — рассмотреть, как эти принципы связаны друг с другом. Именно, допустим, что для действительных чисел имеет место один из принципов непрерывности, и исследуем, какие из остальных принципов непрерывности вытекают отсюда.

Общий вывод, к которому мы придем, это — что все принципы непрерывности эквивалентны.

Будем говорить, что число b является верхней гранью множества E

$$b = \sup E,$$

если 1) $x \leqslant b$ для любого $x \in E$ и 2) не существует числа $b' < b$, обладающего тем же свойством.

Покажем, что из принципа Дедекинда вытекает следующее предложение: каждое непустое ограниченное сверху множество чисел E имеет верхнюю грань. В самом деле, разобьем все действительные числа на два класса X и Y по следующему признаку: положим $x \in X$, если существует такое $a \in E$, что $a \geqslant x$, и положим $y \in Y$, если для любого $a \in E$ имеем $a < y$. Легко проверить, что это — сечение. Согласно принципу Дедекинда, оно имеет рубеж ξ; этот рубеж и будет верхней гранью множества E.

Покажем теперь, что из принципа Дедекинда вытекает принцип Вейерштрасса. Пусть $\{x_n\}$ — неубывающая ограниченная сверху последовательность действительных чисел. По только что доказанному она имеет верхнюю грань ξ. Согласно определению верхней грани $x_n \leqslant \xi$ $(n = 1, 2, \ldots)$, для любого $\varepsilon > 0$ найдется номер n_0, такой, что $x_{n_0} > \xi - \varepsilon$. В силу монотонности последовательности $\{x_n\}$ отсюда вытекает, что $\xi - \varepsilon < x_n \leqslant \xi$ для всех $n > n_0$, т. е. последовательность $\{x_n\}$ сходится и имеет предел ξ.

Для доказательства обратного соотношения между принципами Дедекинда и Вейерштрасса отметим, что из принципа Вейерштрасса вытекает так называемый

Принцип Архимеда. Каковы бы ни были действительные числа $a > 0$ и b, можно найти такое натуральное число n, что $na > b$.

Этот принцип означает, что для любого действительного числа b последовательность $\left\{\frac{b}{n}\right\}$ $(n = 1, 2, \ldots)$ сходится к нулю.

Допустим, что выполняется принцип Вейерштрасса и не выполняется принцип Архимеда. Последнее означает, что существует такое

$a>0$, что последовательность $x_n=na$ ограничена. Кроме того, она возрастающая. По принципу Вейерштрасса она имеет некоторый предел ξ. Отсюда вытекает, что отрезок $\left[\xi-\frac{a}{2}, \xi\right]$ содержит некоторую точку $x_n=na$ нашей последовательности. Но тогда $x_{n+1}=(n+1)a>\xi$, что противоречит тому, что ξ есть верхняя грань $\{x_n\}$.

Из принципа Вейерштрасса вытекает принцип Дедекинда. Пусть множество всех действительных чисел разбито на сумму двух непересекающихся множеств X и Y, так что $x<y$ для любых $x\in X$, $y\in Y$. Покажем, что это сечение имеет единственный рубеж ξ. Пусть m — целое, n — натуральное. Обозначим через x_n наибольший элемент вида $\frac{m}{2^n}\in X$, так что $x_n+\frac{1}{2^n}\in Y$. Так как множество элементов вида $\frac{m}{2^n}$ содержится в множестве элементов вида $\frac{m}{2^{n+1}}$, то $x_n\leqslant x_{n+1}$. Кроме того, последовательность $\{x_n\}$ ограничена (например, числом $x_1+\frac{1}{2}$). Отсюда по принципу Вейерштрасса она имеет некоторый предел ξ. Докажем, что ξ и есть рубеж нашего сечения. Действительно, если $x<\xi$, то $x\in X$. Если же $y>\xi$, то $y\in Y$, так как из принципа Архимеда вытекает, что найдется такой номер n, что $\frac{1}{2^n}<y-\xi=a$. Но $x_n<\xi$, $x_n+\frac{1}{2^n}\in Y$, а тогда $y=\xi+a>x_n+\frac{1}{2^n}$ и, следовательно, $y\in Y$.

Можно также установить, что принципы Кантора и Коши эквивалентны. Однако из выполнения, например, принципа Коши еще не вытекает, что выполняется принцип Дедекинда. Это утверждение надо понимать в следующем смысле: существует упорядоченное поле, для которого принцип Коши выполняется, а принцип Дедекинда не выполняется. Если же заранее предположить, что выполняется принцип Архимеда, то все четыре принципа эквивалентны.

Несчетность континуума. Покажем, что множество всех точек отрезка $0\leqslant x\leqslant 1$ несчетно. Докажем это предложение от противного. Предположим, что множество всех точек отрезка $0\leqslant x\leqslant 1$ счетно. Тогда все точки x этого отрезка можно занумеровать при помощи натуральных чисел

$$x_1, x_2, \ldots, x_n, \ldots \tag{1}$$

Выберем на отрезке $[0, 1]$ отрезок σ_1 так, чтобы его длина была меньше, чем 1, и чтобы он не содержал точки x_1. Такой отрезок наверняка найдется. Далее, внутри отрезка σ_1 выберем отрезок σ_2 так, чтобы длина σ_2 была меньше, чем $\frac{1}{2}$, и чтобы отрезок σ_2 не содержал точек x_1 и x_2. Вообще, после того как выбран отрезок σ_{n-1}, мы выбираем в нем отрезок σ_n так, чтобы длина его была меньше, чем $\frac{1}{n}$, и чтобы

отрезок σ_n не содержал точек $x_1, x_2, \ldots, x_n$. Мы построим таким образом бесконечную последовательность отрезков

$$\sigma_1, \sigma_2, \ldots, \sigma_n \ldots,$$

такую, что каждый последующий отрезок содержится в предшествующем и длины отрезков стремятся к нулю с возрастанием n. Тогда в силу принципа Кантора (см. стр. 15) существует единственная точка x отрезка $[0, 1]$, принадлежащая всем отрезкам σ_n. Так как, по нашему предположению, все точки отрезка $[0, 1]$ записаны в ряду (1), то и точка x, общая всем отрезкам σ_n, совпадает с какой-то точкой x_m этого ряда. Но по построению уже отрезок σ_m не содержит точки x_m, и, следовательно, $x \neq x_m$. Таким образом, мы пришли к противоречию. Поэтому исходное предположение о счетности множества всех точек отрезка $[0, 1]$ ложно, и, значит, множество всех точек отрезка $[0, 1]$ несчетно, что и требовалось доказать.

Эта теорема показывает, что существуют различные бесконечные мощности, и, следовательно, дает положительный ответ на первый из вопросов, поставленных на стр. 8.

§ 4. ТОЧЕЧНЫЕ МНОЖЕСТВА

В предыдущем параграфе мы уже встречались со множествами, элементами которых являются *точки*. В частности, мы рассматривали множество всех точек какого-либо отрезка, множество всех точек (x, y) квадрата $0 \leqslant x \leqslant 1$, $0 \leqslant y \leqslant 1$. Теперь мы займемся более подробно изучением свойств таких множеств.

Множества, элементами которых являются точки, называются *точечными множествами*. Таким образом, можно говорить о точечных множествах на прямой, на плоскости, в каком-либо пространстве. Ради простоты мы ограничимся рассмотрением точечных множеств *на прямой*.

Между действительными числами и точками на прямой имеется тесная связь: каждому действительному числу можно отнести точку на прямой и обратно. Поэтому, говоря о точечных множествах, мы будем причислять к ним и множества, состоящие из действительных чисел — множества на числовой прямой. Обратно: для того чтобы задать точечное множество на прямой, мы будем обычно задавать координаты всех точек нашего множества.

Точечные множества (и, в частности, точечные множества на прямой) обладают рядом особых свойств, отличающих их от произвольных множеств и выделяющих теорию точечных множеств в самостоятельную математическую дисциплину. Прежде всего имеет смысл говорить о *расстоянии* между двумя точками. Далее, между точками на прямой можно установить соотношения *порядка* (левее, правее); в соответствии

с этим говорят, что точечное множество на прямой является *упорядоченным* множеством. Наконец, как уже отмечалось выше, для прямой справедлив принцип Кантора; это свойство прямой принято характеризовать как *полноту* прямой.

Введем обозначения для простейших множеств на прямой.

Отрезок $[a, b]$ — множество точек, координаты которых удовлетворяют неравенствам $a \leqslant x \leqslant b$.

Интервал (a, b) — множество точек, координаты которых удовлетворяют условиям $a < x < b$.

Полуинтервалы $[a, b)$ и $(a, b]$ определяются соответственно условиями: $a \leqslant x < b$ и $a < x \leqslant b$.

Интервалы и полуинтервалы могут быть *несобственными*. Именно, $(-\infty, \infty)$ обозначает всю прямую, а, например, $(-\infty, b]$ — множество всех точек, для которых $x \leqslant b$.

Начнем с рассмотрения различных возможностей расположения множества *в целом* на прямой.

Ограниченные и неограниченные множества. Множество E точек на прямой может либо состоять из точек, расстояния которых от начала координат не превосходят некоторого положительного числа, либо иметь точки, сколь угодно далекие от начала координат. В первом случае множество E называется *ограниченным*, а во втором — *неограниченным*. Примером ограниченного множества может служить множество всех точек отрезка $[0, 1]$, а примером неограниченного множества — множество всех точек с целыми координатами.

Нетрудно видеть, что если a — фиксированная точка на прямой, то множество E будет ограничено в том и только в том случае, если расстояния от точки a до любой точки $x \in E$ не превосходят некоторого положительного числа.

Множества, ограниченные сверху и снизу. Пусть E — множество точек на прямой. Если на прямой существует такая точка A, что любая точка $x \in E$ расположена левее точки A, то говорят, что множество E *ограничено сверху*. Аналогично, если на прямой существует такая точка a, что любая точка $x \in E$ расположена правее точки a, то множество E называется *ограниченным снизу*. Так, множество всех точек на прямой с положительными координатами ограничено снизу, а множество всех точек с отрицательными координатами ограничено сверху.

Ясно, что данное выше определение ограниченного множества эквивалентно следующему: множество E точек на прямой называется ограниченным, если оно ограничено сверху и снизу. Несмотря на то, что эти два определения очень похожи друг на друга, между ними имеется существенное различие: первое основано на том, что между точками на прямой определено расстояние, а второе, что эти точки образуют упорядоченное множество.

Можно также сказать, что множество ограничено, если оно целиком расположено на некотором отрезке $[a, b]$.

Верхняя и нижняя грань множества. Пусть множество E ограничено сверху. Тогда на прямой существуют точки A, правее которых нет ни одной точки множества E. Используя принцип Кантора, можно показать, что среди всех точек A, обладающих этим свойством, найдется самая левая. Эта точка называется *верхней гранью* множества E. Аналогично определяется *нижняя грань* точечного множества.

Если во множестве E есть самая правая точка, то она, очевидно, и будет верхней гранью множества E. Однако может случиться, что во множестве E нет самой правой точки. Например, множество точек с координатами

$$\frac{0}{1}, \frac{1}{2}, \frac{2}{3}, \frac{3}{4}, \frac{4}{5}, \ldots$$

ограничено сверху и не имеет самой правой точки. В таком случае верхняя грань a не принадлежит множеству E, но сколь угодно близко к a имеются точки множества E. В приведенном выше примере $a=1$.

Расположение точечного множества вблизи какой-либо точки на прямой. Пусть E — точечное множество и x — какая-либо точка на прямой. Рассмотрим различные возможности расположения множества E вблизи точки x. Возможны следующие случаи:

1. Ни точка x, ни достаточно близкие к ней точки не принадлежат множеству E.

2. Точка x не принадлежит E, но сколь угодно близко к ней имеются точки множества E.

3. Точка x принадлежит E, но все достаточно близкие к ней точки не принадлежат E.

4. Точка x принадлежит E, и сколь угодно близко к ней имеются другие точки множества E.

В случае 1 точка x называется *внешней* к множеству E, в случае 3 — *изолированной* точкой множества E, а в случаях 2 и 4 — *предельной точкой* множества E.

Таким образом, если $x \overline{\in} E$, то точка x может быть либо внешней к E, либо предельной для него, а если $x \in E$, то она может быть либо изолированной точкой множества E, либо его предельной точкой.

Предельная точка может принадлежать и не принадлежать множеству E и характеризуется тем условием, что сколь угодно близко к ней имеются точки множества E. Иными словами, точка x является предельной точкой множества E, если любой интервал δ, содержащий точку x, содержит бесконечно много точек множества E. Понятие предельной точки является одним из весьма важных понятий теории точечных множеств.

Если точка x и все достаточно близкие к ней точки принадлежат множеству E, то такая точка x называется *внутренней* точкой E. Всякая точка x, которая не является для E ни внешней, ни внутренней, называется *граничной* точкой множества E.

Укажем несколько примеров, поясняющих все эти понятия.

Пример 1. Пусть множество E_1 состоит из точек с координатами

$$1, \frac{1}{2}, \frac{1}{3}, \ldots, \frac{1}{n}, \ldots$$

Тогда каждая точка этого множества является его изолированной точкой, точка 0 есть предельная точка E_1 (не принадлежащая этому множеству), а все остальные точки на прямой — внешние к E_1.

Пример 2. Пусть множество E_2 состоит из всех рациональных точек отрезка $[0, 1]$. Это множество не имеет изолированных точек, каждая точка отрезка $[0, 1]$ является предельной точкой E_2, а все остальные точки на прямой — внешние к E_2. Ясно, что среди предельных точек множества E_2 имеются как принадлежащие к нему, так и не принадлежащие ему.

Пример 3. Пусть множество E_3 состоит из всех точек отрезка $[0,1]$. Как и в предыдущем примере, множество E_3 не имеет изолированных точек, и каждая точка отрезка $[0, 1]$ является его предельной точкой. Однако, в отличие от предыдущего примера, все предельные точки E_3 принадлежат этому множеству.

Пример 4. Пусть множество E_4 состоит из всех точек с целыми координатами на прямой. Каждая точка E_4 является его изолированной точкой; множество E_4 не имеет предельных точек.

Отметим также, что в примере 3 всякая точка интервала $(0, 1)$ является внутренней точкой E_3, а в примере 2 всякая точка отрезка $[0, 1]$ — граничная точка E_2.

Из приведенных выше примеров видно, что бесконечное множество точек на прямой может иметь изолированные точки (E_1, E_4), а может их не иметь (E_2, E_3); точно так же оно может иметь внутренние точки (E_3) и может их не иметь (E_1, E_2, E_4). Что же касается предельных точек, то лишь множество E_4 примера 4 не имеет ни одной предельной точки. Как показывает следующая важная теорема, это связано с тем, что множество E_4 неограничено.

Теорема Больцано—Вейерштрасса. *Всякое ограниченное бесконечное множество точек на прямой имеет хотя бы одну предельную точку.*

Докажем эту теорему. Пусть E — ограниченное бесконечное множество точек на прямой. Так как множество E ограничено, то оно целиком расположено на некотором отрезке $[a, b]$. Разделим этот отрезок пополам. Так как множество E бесконечно, то хотя бы в одном из полученных отрезков лежит бесконечно много точек множества E. Обозначим этот отрезок через σ_1 (если в обеих половинах

отрезка $[a, b]$ лежит бесконечно много точек множества E, то через σ_1 можно обозначить, например, левую). Далее, разделим отрезок σ_1 на два равных отрезка. Так как часть множества E, расположенная на отрезке σ_1, бесконечна, то хотя бы один из полученных отрезков содержит бесконечно много точек множества E. Обозначим этот отрезок через σ_2. Продолжим неограниченно процесс деления отрезков пополам и будем каждый раз брать ту половину, которая содержит бесконечно много точек множества E. Мы получим последовательность отрезков $\sigma_1, \sigma_2, \ldots, \sigma_n, \ldots$ Эта последовательность отрезков обладает такими свойствами: каждый следующий отрезок σ_{n+1} содержится в предыдущем σ_n; каждый отрезок σ_n содержит бесконечно много точек множества E; длины отрезков σ_n стремятся к нулю. Первые два свойства последовательности непосредственно вытекают из ее построения, а для доказательства последнего свойства достаточно заметить, что если длина отрезка $[a, b]$ равна l, то длина отрезка σ_n равна $\frac{l}{2^n}$.

В силу принципа Кантора существует единственная точка x, принадлежащая всем отрезкам σ_n. Покажем, что эта точка x является предельной точкой множества E. Для этого достаточно установить, что если δ есть некоторый интервал, содержащий точку x, то он содержит бесконечно много точек множества E. Так как каждый отрезок σ_n содержит точку x и длины отрезков σ_n стремятся к нулю, то при достаточно большом n отрезок σ_n будет целиком содержаться в интервале δ. Но по условию σ_n содержит бесконечно много точек множества E. Поэтому и δ содержит бесконечно много точек множества E. Итак, точка x действительно является предельной точкой множества E, и теорема доказана.

Упражнение. Покажите, что если множество E ограничено сверху и не имеет самой правой точки, то его верхняя грань является предельной точкой E (и не принадлежит E).

Замкнутые и открытые множества. Одна из основных задач теории точечных множеств — изучение свойств различных типов точечных множеств. Мы познакомим читателя с этой теорией на двух примерах. Именно, мы изучим здесь свойства так называемых замкнутых и открытых множеств.

Множество называется *замкнутым*, если оно содержит все свои предельные точки. Если множество не имеет ни одной предельной точки, то его тоже принято считать замкнутым. Кроме своих предельных точек, замкнутое множество может также содержать изолированные точки. Множество называется *открытым*, если каждая его точка является для него внутренней.

Приведем примеры замкнутых и открытых множеств. Всякий отрезок $[a, b]$ есть замкнутое множество, а всякий интервал (a, b) — открытое множество. Несобственные полуинтервалы $(-\infty, b]$ и $[a, \infty)$

замкнуты, а несобственные интервалы $(-\infty, b)$ и (a, ∞) открыты. Вся прямая является одновременно и замкнутым и открытым множеством. Удобно считать пустое множество тоже одновременно замкнутым и открытым. Любое конечное множество точек на прямой замкнуто, так как оно не имеет предельных точек. Множество, состоящее из точек

$$0, 1, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \ldots \frac{1}{n}, \ldots,$$

замкнуто; это множество имеет единственную предельную точку $x=0$, которая принадлежит множеству.

Наша задача состоит в том, чтобы выяснить, как устроено произвольное замкнутое или открытое множество. Для этого нам понадобится ряд вспомогательных фактов, которые мы примем без доказательства.

1. Пересечение любого числа замкнутых множеств замкнуто.

2. Сумма любого числа открытых множеств есть открытое множество.

3. Если замкнутое множество ограничено сверху, то оно содержит свою верхнюю грань. Аналогично, если замкнутое множество ограничено снизу, то оно содержит свою нижнюю грань.

Пусть E — произвольное множество точек на прямой. Назовем *дополнением* множества E и обозначим через CE множество всех точек на прямой, не принадлежащих множеству E. Ясно, что если x есть внешняя точка для E, то она является внутренней точкой для множества CE и обратно.

4. Если множество F замкнуто, то его дополнение CF открыто и обратно.

Предложение 4 показывает, что между замкнутыми и открытыми множествами имеется весьма тесная связь: одни являются дополнениями других. В силу этого достаточно изучить одни замкнутые или одни открытые множества. Знание свойств множеств одного типа позволяет сразу выяснить свойства множеств другого типа. Например, всякое открытое множество получается путем удаления из прямой некоторого замкнутого множества.

Приступаем к изучению свойств замкнутых множеств. Введем одно определение. Пусть F — замкнутое множество. Интервал (a, b), обладающий тем свойством, что ни одна из его точек не принадлежит множеству F, а точки a и b принадлежат F, называется *смежным интервалом* множества F. К числу смежных интервалов мы будем также относить несобственные интервалы (a, ∞) или $(-\infty, b)$, если точка a или точка b принадлежит множеству F, а сами интервалы с F не пересекаются. Покажем, что если точка x не принадлежит замкнутому множеству F, то она принадлежит одному из его смежных интервалов.

Обозначим через F_x часть множества F, расположенную правее точки x. Так как сама точка x не принадлежит множеству F, то F_x можно представить в форме пересечения

$$F_x = F \cdot [x, \infty).$$

Каждое из множеств F и $[x, \infty)$ замкнуто. Поэтому, в силу предложения 1, множество F_x замкнуто. Если множество F_x пусто, то весь полуинтервал $[x, \infty)$ не принадлежит множеству F. Допустим теперь, что множество F_x не пусто. Так как это множество целиком расположено на полуинтервале $[x, \infty)$, то оно ограничено снизу. Обозначим через b его нижнюю грань. Согласно предложению 3, $b \in F_x$, а значит $b \in F$. Далее, так как b есть нижняя грань множества F_x, то полуинтервал $[x, b)$, лежащий левее точки b, не содержит точек множества F_x и, следовательно, не содержит точек множества F. Итак, мы построили полуинтервал (x, b), не содержащий точек множества F, причем либо $b = \infty$, либо точка b принадлежит множеству F. Аналогично строится полуинтервал $(a, x]$, не содержащий точек множества F, причем либо $a = -\infty$, либо $a \in F$. Теперь ясно, что интервал (a, b) содержит точку x и является смежным интервалом множества F. Легко видеть, что если (a_1, b_1) и (a_2, b_2) — два смежных интервала множества F, то эти интервалы либо совпадают, либо не пересекаются.

Из предыдущего следует, что всякое замкнутое множество на прямой получается путем удаления из прямой некоторого числа интервалов, а именно смежных интервалов множества F. Так как каждый интервал содержит по крайней мере одну рациональную точку, а всех рациональных точек на прямой — счетное множество, то легко убедиться, что число всех смежных интервалов не более чем счетно. Отсюда получаем окончательный вывод. Всякое замкнутое множество на прямой получается путем удаления из прямой не более чем счетного множества непересекающихся интервалов.

В силу предложения 4, отсюда сразу вытекает, что всякое открытое множество на прямой представляет собой не более чем счетную сумму непересекающихся интервалов. В силу предложений 1 и 2, ясно также, что всякое множество, устроенное, как указано выше, действительно является замкнутым (открытым).

Как видно из нижеследующего примера, замкнутые множества могут иметь весьма сложное строение.

Канторово совершенное множество. Построим одно специальное замкнутое множество, обладающее рядом замечательных свойств. Прежде всего удалим из прямой несобственные интервалы $(-\infty, 0)$ и $(1, \infty)$. После этой операции у нас останется отрезок $[0, 1]$. Далее, удалим из этого отрезка интервал $\left(\frac{1}{3}, \frac{2}{3}\right)$, составляющий его среднюю треть.

Из каждого из оставшихся двух отрезков $\left[0, \frac{1}{3}\right]$ и $\left[\frac{2}{3}, 1\right]$ удалим его среднюю треть. Этот процесс удаления средних третей у остающихся отрезков продолжим неограниченно. Множество точек на прямой, остающееся после удаления всех этих интервалов, называется канторовым совершенным множеством; мы будем обозначать его буквой P.

Рассмотрим некоторые свойства этого множества. Множество P замкнуто, так как оно образуется путем удаления из прямой некоторого множества непересекающихся интервалов. Множество P не пусто; во всяком случае в нем содержатся концы всех выброшенных интервалов.

Замкнутое множество F называется *совершенным*, если оно не содержит изолированных точек, т. е. если каждая его точка является предельной точкой. Покажем, что множество P совершенно. Действительно, если бы некоторая точка x была изолированной точкой множества P, то она служила бы общим концом двух смежных интервалов этого множества. Но, согласно построению, смежные интервалы множества P не имеют общих концов.

Множество P не содержит ни одного интервала. В самом деле, допустим, что некоторый интервал δ целиком принадлежит множеству P. Тогда он целиком принадлежит одному из отрезков, получающихся на n-м шаге построения множества P. Но это невозможно, так как при $n \to \infty$ длины этих отрезков стремятся к нулю.

Можно показать, что множество P имеет мощность континуума. В частности, отсюда следует, что канторово совершенное множество содержит, кроме концов смежных интервалов, еще и другие точки. Действительно, концы смежных интервалов образуют лишь счетное множество.

Разнообразные типы точечных множеств постоянно встречаются в самых различных разделах математики, и знание их свойств совершенно необходимо при исследовании многих математических проблем. Особенно большое значение имеет теория точечных множеств для математического анализа и топологии.

Приведем несколько примеров появления точечных множеств в классических разделах анализа. Пусть $f(x)$ — непрерывная функция, заданная на отрезке $[a, b]$. Зафиксируем число α и рассмотрим множество тех точек x, для которых $f(x) \geqslant \alpha$. Нетрудно показать, что это множество может быть произвольным замкнутым множеством, расположенным на отрезке $[a, b]$. Точно так же множество точек x, для которых $f(x) > \alpha$, может быть каким угодно открытым множеством $G \subset [a, b]$. Если $f_1(x), f_2(x), \ldots, f_n(x); \ldots$ есть последовательность непрерывных функций, заданных на отрезке $[a, b]$, то множество тех точек x, где эта последовательность сходится, не может быть произвольным, а принадлежит к вполне определенному типу.

Математическая дисциплина, занимающаяся изучением строения точечных множеств, называется *дескриптивной теорией множеств*. Весьма большие заслуги в деле развития дескриптивной теории множеств принадлежат советским математикам — Н. Н. Лузину и его ученикам П. С. Александрову, М. Я. Суслину, А. Н. Колмогорову, М. А. Лаврентьеву, П. С. Новикову, Л. В. Келдыш, А. А. Ляпунову и др.

Исследования Н. Н. Лузина и его учеников показали, что имеется глубокая связь между дескриптивной теорией множеств и математической логикой. Трудности, возникающие при рассмотрении ряда задач дескриптивной теории множеств (в частности, задач об определении мощности тех или иных множеств), являются трудностями логической природы. Напротив, методы математической логики позволяют более глубоко проникнуть в некоторые вопросы дескриптивной теории множеств.

§ 5. МЕРА МНОЖЕСТВ

Понятие меры множества является далеко идущим обобщением понятия длины отрезка. В простейшем случае (которым только мы и будем заниматься) задача состоит в том, чтобы дать определение длины не только для отрезков, но также и для более сложных точечных множеств, расположенных на прямой.

Примем за единицу измерения отрезок $[0, 1]$. Тогда длина произвольного отрезка $[a, b]$, очевидно, равна $b-a$. Точно так же если имеется два непересекающихся отрезка $[a_1, b_1]$ и $[a_2, b_2]$, то под длиной множества E, состоящего из этих двух отрезков, естественно понимать число $(b_1-a_1)+(b_2-a_2)$. Однако далеко не так ясно, что́ следует понимать под длиной множества более сложной природы, расположенного на прямой; например, чему равна длина канторова множества P, рассмотренного в § 4 этой главы? Отсюда вывод: понятие длины множества, расположенного на прямой, нуждается в строгом математическом определении.

Задача определения длины множеств, или, как говорят еще, задача измерения множеств, весьма важна, так как она имеет существенное значение для обобщения понятия интеграла. Понятие меры множества применяется и в других вопросах теории функций, а также в теории вероятностей, топологии, функциональном анализе и т. д.

Ниже излагается определение меры множеств, предложенное французским математиком А. Лебегом и лежащее в основе данного им определения интеграла.

Мера открытого и замкнутого множества. Начнем с определения меры произвольного открытого или замкнутого множества. Как уже отмечалось в § 4, всякое открытое множество на прямой является конечной или счетной суммой попарно не пересекающихся интервалов.

Мерой открытого множества называется сумма длин составляющих его интервалов.

Таким образом, если

$$G=\sum(a_i,\ b_i)$$

и интервалы $(a_i,\ b_i)$ попарно не пересекаются, то мера G равна $\sum(b_i-a_i)$. Обозначая вообще меру множества E через μE, можем написать

$$\mu G=\sum(b_i-a_i).$$

В частности, мера одного интервала равна его длине

$$\mu(a,\ b)=b-a.$$

Всякое замкнутое множество F, содержащееся в отрезке $[a,\ b]$ и такое, что концы отрезка $[a,\ b]$ принадлежат F, получается из отрезка $[a,\ b]$ путем удаления из него некоторого открытого множества G. В соответствии с этим *мерой замкнутого множества* $F\subseteq[a,\ b]$, где $a\in F$, $b\in F$, называется разность между длиной отрезка $[a,\ b]$ и мерой открытого множества G, дополнительного к F (относительно $[a,\ b]$).

Итак,

$$\mu F=(b-a)-\mu G. \tag{2}$$

Нетрудно усмотреть, что, согласно этому определению, мера произвольного отрезка равна его длине

$$\mu[a,\ b]=b-a,$$

а мера множества, состоящего из конечного числа точек, равна нулю.

Общее определение меры. Для того чтобы дать определение меры множеств более общей природы, чем открытые и замкнутые, нам понадобится одно вспомогательное понятие. Пусть E — некоторое множество, лежащее на отрезке $[a,\ b]$. Рассмотрим всевозможные *покрытия* множества E, т. е. всевозможные открытые множества $V(E)$, содержащие E. Мера каждого из множеств $V(E)$ уже определена. Совокупность мер всех множеств $V(E)$ есть некоторое множество положительных чисел. Это множество чисел ограничено снизу (хотя бы числом 0) и потому имеет нижнюю грань, которую мы обозначим через $\mu_e E$. Число $\mu_e E$ называется *внешней мерой* множества E.

Пусть $\mu_e E$ — внешняя мера множества E, а $\mu_e CE$ — внешняя мера его дополнения относительно отрезка $[a,\ b]$.

Если удовлетворяется соотношение

$$\mu_e E+\mu_e CE=b-a, \tag{3}$$

то множество E называется *измеримым*, а число $\mu_e E$ — его *мерой*: $\mu E=\mu_e E$; если соотношение (3) не удовлетворяется, то говорят, что множество E неизмеримо; неизмеримое множество не имеет меры.

Отметим, что всегда

$$\mu_e E + \mu_e CE \geqslant b - a. \tag{4}$$

Сделаем несколько пояснений. Длина простейших множеств (например, интервалов и отрезков) обладает рядом замечательных свойств. Укажем важнейшие из них.

1. Если множества E_1 и E измеримы и $E_1 \subseteq E$, то

$$\mu E_1 \leqslant \mu E,$$

т. е. мера части множества E не превосходит меры всего множества E.

2. Если множества E_1 и E_2 измеримы, то множество $E = E_1 + E_2$ измеримо и

$$\mu(E_1 + E_2) \leqslant \mu E_1 + \mu E_2,$$

т. е. мера суммы не превосходит суммы мер слагаемых.

3. Если множества $E_i\,(i = 1, 2, \ldots)$ измеримы и попарно не пересекаются, $E_i E_j = 0\,(i \neq j)$, то их сумма $E = \sum E_i$ измерима и

$$\mu(\sum E_i) = \sum \mu E_i,$$

т. е. мера конечной или счетной суммы попарно непересекающихся множеств равна сумме мер слагаемых.

Это свойство меры называется ее полной аддитивностью.

4. Мера множества E не меняется, если его сдвинуть как твердое тело.

Желательно, чтобы основные свойства длины сохранялись и для более общего понятия меры множеств. Но, как можно совершенно строго показать, это оказывается *невозможным*, если приписывать меру произвольному множеству точек на прямой. Поэтому-то в данном выше определении и появляются множества, имеющие меру или измеримые, и множества, не имеющие меры или неизмеримые. Впрочем, класс измеримых множеств настолько широк, что это обстоятельство не вносит каких-либо существенных неудобств. Даже построение примера неизмеримого множества представляет известные трудности.

Приведем несколько примеров измеримых множеств.

Пример 1. *Мера канторова совершенного множества* P (см. § 4). При построении множества P из отрезка $[0, 1]$ выбрасывается сперва один смежный интервал длины $1/3$, затем два смежных интервала длины $1/9$, затем четыре смежных интервала длины $1/27$ и т. д. Вообще, на n-м шаге выбрасывается 2^{n-1} смежных интервалов длины $\frac{1}{3^n}$. Таким образом, сумма длин всех выброшенных интервалов равна

$$S = \frac{1}{3} + \frac{2}{9} + \frac{4}{27} + \ldots + \frac{2^{n-1}}{3^n} + \ldots$$

Члены этого ряда представляют собою геометрическую прогрессию с первым членом $^1/_3$ и знаменателем $^2/_3$. Поэтому сумма ряда S равна

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

Итак, сумма длин всех смежных к канторовому множеству интервалов равна 1. Иначе говоря, мера дополнительного к P открытого множества G равна 1. Поэтому само множество имеет меру

$$\mu P=1-\mu G=1-1=0.$$

Как показывает этот пример, множество может иметь мощность континуума и тем не менее иметь меру, равную нулю.

Пример 2. *Мера множества R всех рациональных точек отрезка* [0, 1]. Покажем прежде всего, что $\mu_e R=0$. В § 2 было установлено, что множество R счетно. Расположим точки множества R в последовательность

$$r_1, r_2, \ldots, r_n, \ldots$$

Далее, зададим $\varepsilon>0$ и окружим точку r_n интервалом δ_n длины $\frac{\varepsilon}{2^n}$. Сумма $\delta=\sum\delta_n$ есть открытое множество, покрывающее R. Интервалы δ_n могут пересекаться, поэтому

$$\mu(\delta)=\mu\left(\sum\delta_n\right)\leqslant\sum\mu\delta_n=\sum\frac{\varepsilon}{2^n}=\varepsilon.$$

Так как ε можно выбрать сколь угодно малым, то $\mu_e R=0$.

Далее, согласно (3)

$$\mu_e R+\mu_e CR\geqslant 1,$$

т. е. $\mu_e CR\geqslant 1$. Так как CR содержится в отрезке [0, 1], то $\mu_e CR\leqslant 1$. Итак,

$$\mu_e R+\mu_e CR=1,$$

откуда

$$\mu R=0,\quad \mu CR=1\ ^1. \tag{5}$$

Этот пример показывает, что множество может быть всюду плотным на некотором отрезке и тем не менее иметь меру, равную нулю.

Множества меры нуль во многих вопросах теории функций не играют никакой роли, и ими следует пренебрегать. Например, функция $f(x)$ интегрируема по Риману в том и только в том случае, если она ограничена и множество ее точек разрыва имеет меру нуль. Можно было бы привести значительное число таких примеров.

[1] То же самое рассуждение показывает, что всякое счетное множество точек на прямой имеет меру нуль.

Измеримые функции. Переходим к одному из наиболее блестящих приложений понятия меры множеств, а именно к описанию того класса функций, с которыми фактически оперирует математический анализ и теория функций. Точная постановка задачи такова. Если последовательность функций $\{f_n(x)\}$, заданных на некотором множестве E, сходится в каждой точке E, кроме, быть может, точек множества N меры нуль, то будем говорить, что последовательность $\{f_n(x)\}$ сходится *почти всюду*.

Какие функции можно получить из непрерывных функций путем повторного применения операции построения предела почти всюду сходящейся последовательности функций и алгебраических операций?

Для ответа на этот вопрос нам потребуется несколько новых понятий.

Пусть функция $f(x)$ определена на некотором множестве E и α — произвольное действительное число. Обозначим через

$$E[f(x)>\alpha]$$

множество тех точек E, для которых $f(x)>\alpha$. Например, если функция $f(x)$ определена на отрезке $[0, 1]$ и на этом отрезке $f(x)=x$, то множества $E[f(x)>\alpha]$ равны $[0, 1]$ для $\alpha<0$, равны $(\alpha, 1]$ для $0\leqslant\alpha<1$ и пусты для $\alpha\geqslant 1$.

Функция $f(x)$, определенная на некотором множестве E, называется измеримой, если само множество E измеримо и для любого действительного числа α измеримо множество $E[f(x)>\alpha]$.

Можно показать, что произвольная непрерывная функция, заданная на отрезке, измерима. Однако к числу измеримых функций принадлежат также и многие разрывные функции, например функция Дирихле, равная 1 для иррациональных точек отрезка $[0, 1]$ и равная 0 для остальных точек этого отрезка.

Отметим без доказательства, что измеримые функции обладают следующими свойствами.

1. Если $f(x)$ и $\varphi(x)$ — измеримые функции, определенные на одном и том же множестве E, то функции

$$f+\varphi,\quad f-\varphi,\quad f\cdot\varphi \quad \text{и} \quad \frac{f}{\varphi}$$

также измеримы (последняя, если $\varphi\neq 0$).

Это свойство показывает, что алгебраические операции над измеримыми функциями снова приводят к измеримым функциям.

2. Если последовательность измеримых функций $\{f_n(x)\}$, определенных на множестве E, сходится почти всюду к функции $f(x)$, то эта функция также измерима.

Таким образом, операция построения предела почти всюду сходящейся последовательности измеримых функций вновь приводит к измеримым функциям.

Эти свойства измеримых функций были установлены Лебегом. Глубокое исследование измеримых функций было произведено советскими математиками Д. Ф. Егоровым и Н. Н. Лузиным. В частности, Н. Н. Лузин показал, что всякую измеримую функцию, заданную на отрезке, можно превратить в непрерывную, изменив ее значения на некотором множестве сколь угодно малой меры.

Этот классический результат Н. Н. Лузина и перечисленные выше свойства измеримых функций позволяют показать, что измеримые функции и представляют собой тот класс функций, о котором шла речь в начале этого пункта. Измеримые функции имеют также большое значение для теории интегрирования, именно, понятие интеграла может быть обобщено таким образом, чтобы всякая ограниченная измеримая функция оказалась интегрируемой. Подробнее об этом рассказывается в следующем параграфе.

§ 6. ИНТЕГРАЛ ЛЕБЕГА

Переходим к центральному вопросу этой главы — определению и описанию свойств интеграла Лебега.

Чтобы понять принцип устройства этого интеграла, рассмотрим следующий пример. Пусть имеется большое количество монет различного достоинства и требуется сосчитать общую сумму денег, заключенную в этих монетах. Это можно сделать двумя способами. Можно откладывать монеты подряд и прибавлять стоимость каждой новой монеты к общей стоимости всех ранее отложенных. Однако можно поступить и иначе: сложить монеты стопочками так, чтобы в каждой стопочке были монеты одного достоинства, затем сосчитать число монет в каждой стопочке, умножить это число на стоимость соответствующей монеты, а затем сложить полученные числа. Первый способ счета денег соответствует процессу интегрирования Римана, а второй — процессу интегрирования Лебега.

Переходя от монет к функциям, мы можем сказать, что для вычисления интеграла Римана производится деление на мелкие части области задания функции (оси абсцисс; рис. 2а), а для вычисления интеграла Лебега производится деление области значений функции (оси ординат; рис. 2б). Последний принцип применялся практически задолго до Лебега при вычислении интегралов от функций, имеющих колебательный характер, однако Лебег впервые развил его во всей общности и дал его строгое обоснование при помощи теории меры.

Рассмотрим, как связаны между собою мера множеств и интеграл Лебега. Пусть E — какое-либо измеримое множество, расположенное на некотором отрезке $[a, b]$. Построим функцию $\varphi(x)$, равную 1 для x,

принадлежащих E, и равную 0 для x, не принадлежащих E. Иными словами, положим

$$\varphi(x) = \begin{cases} 1 & x \in E. \\ 0 & x \overline{\in} E. \end{cases}$$

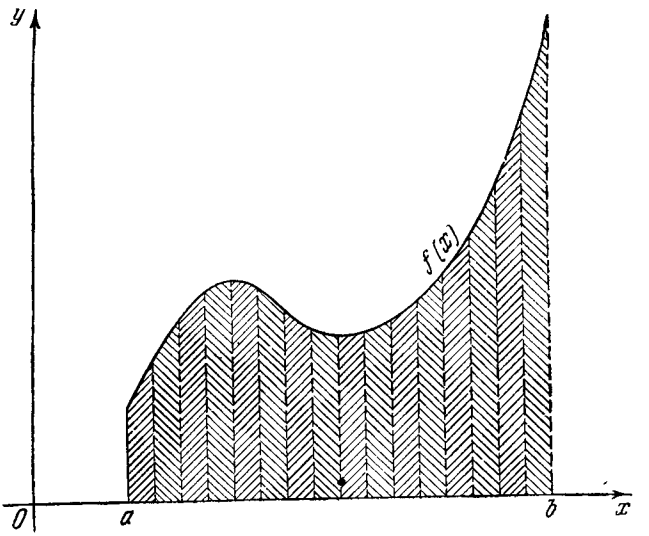


Рис. 2а.

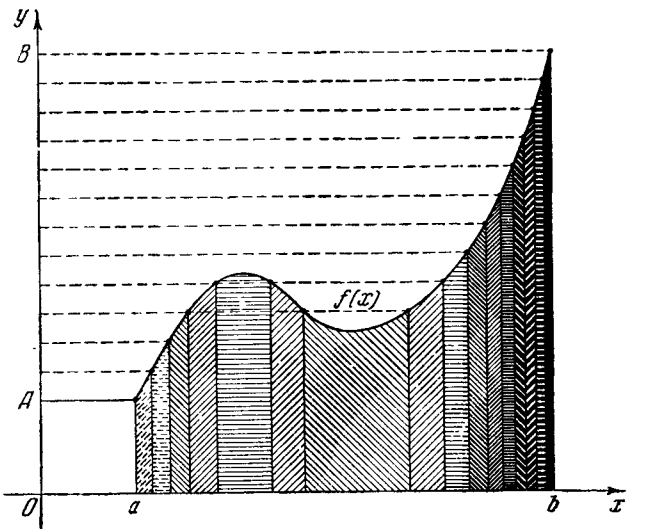


Рис. 2б.

Функцию $\varphi(x)$ принято называть характеристической функцией множества E. Рассмотрим интеграл

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

Мы уже привыкли считать, что интеграл равен площади фигуры D, ограниченной осью абсцисс, прямыми $x=a$, $x=b$ и кривой $y=\varphi(x)$ [см. главу II (том 1)]. Так как в данном случае «высота» фигуры D отлична от нуля и равна 1 для точек $x\in E$ и только для этих точек, то (согласно формуле, площадь равна длине, умноженной на ширину) ее площадь должна быть численно равна длине (мере) множества E. Итак, I должно быть равно мере множества E

$$I=\mu E. \tag{6}$$

Именно так и определяет Лебег интеграл от функции $\varphi(x)$.

Читатель должен твердо уяснить себе, что равенство (6) является *определением* интеграла $\int_a^b \varphi(x)\,dx$ как интеграла Лебега. Может случиться, что интеграл I не будет существовать в том смысле, как это понималось в главе II (том 1), т. е. как предел интегральных сумм. Даже если это последнее имеет место, интеграл I как интеграл Лебега существует и равен μE.

В качестве примера подсчитаем интеграл от функции Дирихле $\Phi(x)$, равной 0 в рациональных точках отрезка [0, 1] и равной 1 в иррациональных точках этого отрезка. Так как, согласно (5), мера множества иррациональных точек отрезка [0, 1] равна 1, то интеграл Лебега

$$\int_0^1 \Phi(x)\,dx$$

равен 1. Нетрудно проверить, что интеграл Римана от этой функции не существует.

Одно вспомогательное предложение. Пусть теперь $f(x)$ — произвольная ограниченная измеримая функция, заданная на отрезке $[a, b]$. Покажем, что всякую такую функцию можно сколь угодно точно представить в виде линейной комбинации характеристических функций множеств. Чтобы убедиться в этом, разобьем отрезок оси ординат между нижней и верхней гранями значений функции A и B точками $y_0=A$, y_1, ..., $y_n=B$ на отрезки длины меньшей ε, где ε — произвольное фиксированное положительное число. Далее, если в точке $x\in[a, b]$

$$y_i\leqslant f(x)<y_{i+1} \quad (i=0, 1, \ldots, n-1),$$

то положим в этой точке

$$\varphi(x)=y_i,$$

а если в точке x

$$f(x)=y_n=B,$$

то положим

$$\varphi(x)=y_n.$$

Построение функции $\varphi(x)$ показано на рис. 3.

Согласно построению функции $\varphi(x)$, в любой точке отрезка $[a, b]$

$$|f(x)-\varphi(x)|<\varepsilon.$$

Кроме того, так как функция $\varphi(x)$ принимает лишь конечное число значений $y_0, y_1, \ldots, y_n$, то ее можно записать в виде

$$\varphi(x)=y_0\cdot\varphi_0(x)+y_1\cdot\varphi_1(x)+\ldots+y_n\cdot\varphi_n(x), \qquad (7)$$

где $\varphi_i(x)$ — характеристическая функция того множества, где $\varphi(x)=y_i$, т. е. $y_i\leqslant f(x)<y_{i+1}$ (в каждой точке $x\in[a, b]$ лишь одно слагаемое

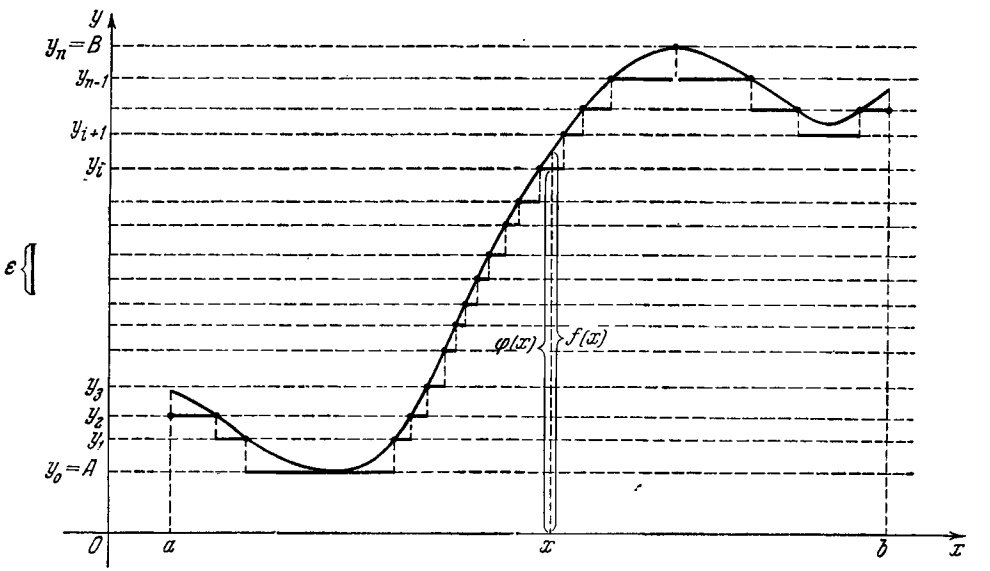


Рис. 3.

в правой части формулы (7) отлично от нуля!) Этим наше предложение установлено.

Определение интеграла Лебега. Переходим к определению интеграла Лебега от произвольной ограниченной измеримой функции. Так как функция $\varphi(x)$ мало отличается от функции $f(x)$, то в качестве приближенного значения интеграла от функции $f(x)$ можно принять интеграл от функции $\varphi(x)$. Но, замечая, что функции $\varphi_i(x)$ являются характеристическими функциями множеств, и пользуясь формально обычными правилами вычисления интеграла, получаем

$$\int_a^b\varphi(x)\,dx=\int_a^b\{y_0\varphi_0(x)+y_1\varphi_1(x)+\ldots+y_n\varphi_n(x)\}\,dx=$$

$$=y_0\int_a^b\varphi_0(x)\,dx+y_1\int_a^b\varphi_1(x)\,dx+\ldots+y_n\int_a^b\varphi_n(x)\,dx=$$

$$=y_0\nu e_0+y_1\nu e_1+\ldots+y_n\nu e_n,$$

где μe_i есть мера множества e_i тех x, для которых

$$y_i \leqslant f(x) < y_{i+1}.$$

Итак, приближенным значением интеграла Лебега от функции $f(x)$ является «интегральная сумма Лебега»

$$S = y_0 \mu e_0 + y_1 \mu e_1 + \ldots + y_n \mu e_n.$$

В соответствии с этим интеграл Лебега определяется как предел интегральных сумм Лебега S, когда

$$\max |y_{i+1} - y_i| \to 0,$$

что соответствует равномерной сходимости функций $\varphi(x)$ к функции $f(x)$.

Можно показать, что интегральные суммы Лебега имеют предел для любой ограниченной измеримой функции, т. е. любая ограниченная измеримая функция интегрируема по Лебегу. Интеграл Лебега можно также распространить на некоторые классы неограниченных измеримых функций, но мы не будем этим заниматься.

Свойства интеграла Лебега. Интеграл Лебега обладает всеми хорошими свойствами обычного интеграла, именно, интеграл от суммы равен сумме интегралов, постоянный множитель можно выносить за знак интеграла и т. д. Однако интеграл Лебега обладает еще одним замечательным свойством, которым обычный интеграл не обладает: если измеримые функции $f_n(x)$ ограничены в совокупности:

$$|f_n(x)| < K$$

для любого n и любого x из отрезка $[a, b]$ и последовательность $\{f_n(x)\}$ сходится почти всюду к функции $f(x)$, то

$$\int_a^b f_n(x)\,dx \to \int_a^b f(x)\,dx.$$

Иными словами, интеграл Лебега допускает безотказный переход к пределу. Именно это свойство интеграла Лебега делает его весьма удобным, а часто и неизбежным инструментом во многих исследованиях. В частности, интеграл Лебега совершенно необходим в теории тригонометрических рядов, в теории функциональных пространств (см. главу XIX) и других разделах математики.

Приведем пример. Пусть $f(x)$ — периодическая функция с периодом 2π и

$$\frac{a_0}{2} + \sum_{n=1}^{\infty} (a_n \cos nx + b_n \sin nx)$$

— ее ряд Фурье. Если, например, функция $f(x)$ непрерывна, то, как нетрудно показать,

$$\frac{1}{\pi}\int_0^{2\pi} f^2(x)\,dx = \frac{a_0^2}{2} + \sum_{n=1}^{\infty}(a_n^2 + b_n^2). \tag{8}$$

Это тождество носит название равенства Парсеваля. Рассмотрим такой вопрос: для какого класса периодических функций справедливо равенство Парсеваля (8)? Ответ на этот вопрос гласит: равенство Парсеваля (8) выполняется в том и только в том случае, если функция $f(x)$ измерима на отрезке $[0, 2\pi]$ и функция $f^2(x)$ интегрируема по Лебегу на этом отрезке.

ЛИТЕРАТУРА

Популярная литература

Лебег А. Об измерении величин. ГОНТИ, 1938.

Маркушевич А. И. Действительные числа и основные принципы теории пределов. Учпедгиз, 1948.

Университетские учебники

Александров П. С. Введение в общую теорию множеств и функций. Гостехиздат, 1948.

Александров П. С. и Колмогоров А. Н. Введение в теорию функций действительного переменного. Изд. 3-е, ГОНТИ, 1938.

Натансон И. П. Теория функций вещественной переменной. Гостехиздат, 1950.

Специальные монографии

Лебег А. Интегрирование и отыскание примитивных. ГТТИ, 1934.

Хаусдорф Ф. Теория множеств. ГОНТИ, 1937.

Глава XVI

ЛИНЕЙНАЯ АЛГЕБРА

§ 1. ПРЕДМЕТ ЛИНЕЙНОЙ АЛГЕБРЫ И ЕЕ АППАРАТ

Линейные функции и матрицы. Среди функций от одной переменной наиболее простой является так называемая *линейная функция* $l(x) = ax + b$. Ее графиком является, как известно, простейшая из линий — прямая.

Вместе с тем линейная функция — одна из важнейших. Дело в том, что всякая «гладкая» линия на малом участке похожа на прямую, и чем меньше участок кривой, тем теснее она примыкает к прямой линии. На языке теории функций это значит, что любая «гладкая» (непрерывно дифференцируемая) функция при малом изменении независимой переменной близка к линейной функции. Линейная функция может быть охарактеризована тем, что ее приращение пропорционально приращению независимой переменной. В самом деле: $\Delta l(x) = l(x_0 + \Delta x) - l(x_0) = a(x_0 + \Delta x) + b - (ax_0 + b) = a\,\Delta x$. Обратно, если $\Delta l(x) = a\,\Delta x$, то $l(x) - l(x_0) = a(x - x_0)$ и $l(x) = ax + l(x_0) - ax_0 = ax + b$, где $b = l(x_0) - ax_0$. Но из дифференциального исчисления мы знаем, что в приращении любой дифференцируемой функции естественным образом выделяется главная часть, так называемый дифференциал функции, пропорциональный приращению независимой переменной, и приращение функции отличается от ее дифференциала на бесконечно малую более высокого порядка, чем приращение независимой переменной. Таким образом дифференцируемая функция при бесконечно малом изменении независимой переменной действительно близка к линейной функции с точностью до бесконечно малой более высокого порядка.

Аналогично обстоит дело и с функциями от нескольких переменных.

Линейной функцией от нескольких переменных называется функция вида $a_1x_1 + a_2x_2 + \ldots + a_nx_n + b$. Если $b = 0$, то линейная функция называется *однородной*. Линейная функция от нескольких переменных характеризуется следующими двумя свойствами:

1. Приращение линейной функции, вычисленное в предположении, что лишь одна из независимых переменных приобретает некоторое приращение при неизменных значениях для остальных, пропорционально приращению этой независимой переменной.

2. Приращение линейной функции, вычисленное в предположении, что все независимые переменные получают некоторое приращение, равно алгебраической сумме приращений, получающихся от изменения каждой переменной в отдельности.

Линейная функция от нескольких переменных играет среди всех функций от этих переменных ту же роль, что и линейная функция от одной переменной среди всех функций одной переменной. Именно, любая «гладкая» функция (т. е. функция, имеющая непрерывные частные производные по всем переменным) при малом изменении независимых переменных близка к некоторой линейной функции. В самом деле, приращение такой функции $w=f(x_1, x_2, \ldots, x_n)$ равно с точностью до малых высших порядков полному дифференциалу $\frac{\partial f}{\partial x_1}dx_1+\ldots+\frac{\partial f}{\partial x_n}dx_n$, который есть линейная однородная функция от приращений независимых переменных $dx_1, \ldots, dx_n$. Отсюда следует, что сама функция w, равная сумме ее начального значения и приращения, выражается через независимые переменные при малом их изменении в виде линейной неоднородной функции с точностью до малых высших порядков.

Задачи, решение которых требует рассмотрения функций от нескольких переменных, возникают в связи с изучением зависимости одной величины от нескольких факторов. Задача называется *линейной*, если оказывается линейной исследуемая зависимость. На основании указанных выше свойств линейной функции, линейная задача может быть охарактеризована следующими свойствами.

1. Свойство пропорциональности. Результат действия каждого отдельного фактора пропорционален его величине.

2. Свойство независимости. Общий результат действия равен сумме результатов действий отдельных факторов.

То, что любая «гладкая» функция при малых изменениях переменных может быть в первом приближении заменена линейной, есть отражение общего принципа — любая задача об изменении некоторой величины в зависимости от действия нескольких факторов может рассматриваться в первом приближении, при малых воздействиях, как задача линейная, т. е. обладающая свойствами независимости и пропорциональности. Часто оказывается, что такое рассмотрение дает удовлетворительный для практики результат (классическая теория упругости, теория малых колебаний и т. д.).

Изучаемые физические величины часто сами характеризуются несколькими числами (сила — тремя проекциями на оси координат, напряженное состояние упругого тела в данной точке — шестью компонентами так называемого тензора напряжения и т. д.). Поэтому возникает необходимость одновременного рассмотрения нескольких функций от нескольких переменных, а в первом приближении — нескольких линейных функций.

Линейная функция от одной переменной настолько проста по своим свойствам, что не требует никакого специального рассмотрения. Иначе обстоит дело с линейными функциями от нескольких переменных, где наличие многих переменных вносит некоторые специфические особенности. Дело еще более осложняется, если перейти от одной функции нескольких переменных $x_1, x_2, \ldots, x_n$ к совокупности нескольких функций $y_1, y_2, \ldots, y_m$ от тех же переменных. Здесь в качестве «первого приближения» появляется совокупность линейных функций:

$$\begin{aligned} y_1 &= a_{11}x_1 + \ldots + a_{1n}x_n + b_1, \\ y_2 &= a_{21}x_1 + \ldots + a_{2n}x_n + b_2, \\ &\ldots\ldots\ldots\ldots\ldots \\ y_m &= a_{m1}x_1 + \ldots + a_{mn}x_n + b_m. \end{aligned}$$

Совокупность линейных функций есть уже довольно сложный математический объект, и его исследование богато интересным и нетривиальным содержанием.

Учение о линейных функциях и их совокупностях и составляет в первую очередь предмет той ветви алгебры, которая называется линейной алгеброй.

Исторически первой задачей линейной алгебры является задача о решении системы линейных уравнений:

$$\begin{aligned} a_{11}x_1 + \ldots + a_{1n}x_n &= b_1, \\ a_{21}x_1 + \ldots + a_{2n}x_n &= b_2, \\ &\ldots\ldots\ldots\ldots \\ a_{m1}x_1 + \ldots + a_{mn}x_n &= b_m. \end{aligned}$$

Простейший случай этой задачи рассматривается в школьном курсе элементарной алгебры. Задача о приемах возможно более простого и наименее трудоемкого численного решения систем при больших n до сих пор привлекает пристальное внимание многих ученых, так как численное решение систем входит как важная составная часть во многие расчеты и исследования.

Линейные однородные функции иначе называются *линейными формами*. Данная система линейных форм

$$\begin{aligned} y_1 &= a_{11}x_1 + \ldots + a_{1n}x_n, \\ &\ldots\ldots\ldots\ldots \\ y_m &= a_{m1}x_1 + \ldots + a_{mn}x_n \end{aligned}$$

описывается системой коэффициентов, так как свойства такой системы форм зависят именно от численных значений коэффициентов, а название переменных не является существенным.

Например, системы форм

$$\begin{matrix} 3x_1 + x_2 - x_3, \\ 2x_1 + x_2 + 3x_3, \\ x_1 - x_2 - x_3 \end{matrix} \quad \text{и} \quad \begin{matrix} 3t_1 + t_2 - t_3, \\ 2t_1 + t_2 + 3t_3, \\ t_1 - t_2 - t_3 \end{matrix}$$

обладают, конечно, одинаковыми свойствами и могут не считаться существенно различными.

Совокупность коэффициентов системы линейных форм естественно задавать в виде прямоугольной таблицы

$$\begin{pmatrix} a_{11} \dots a_{1n} \\ \dots \\ a_{m1} \dots a_{mn} \end{pmatrix}.$$

Такие таблицы носят название *матриц*. Числа a_{ij} называются элементами матрицы. Потребность в рассмотрении матриц с необходимостью возникает из самого предмета линейной алгебры.

Важными частными случаями матриц являются матрицы, состоящие из одного столбца, называемые просто столбцами, матрицы, состоящие из одной строки, называемые строками, и, наконец, квадратные матрицы, т. е. такие, у которых число строк равно числу столбцов. Число строк (или столбцов) квадратной матрицы называется ее *порядком*. «Матрица» (a), состоящая из одного числа a, отождествляется с этим числом.

В связи с простейшими действиями над совокупностью линейных форм естественным образом определяются действия над матрицами.

Пусть даны две системы линейных форм

$$\begin{aligned} y_1 &= a_{11}x_1 + \dots + a_{1n}x_n, \\ &\dots\dots\dots\dots \\ y_m &= a_{m1}x_1 + \dots + a_{mn}x_n, \end{aligned}$$

и

$$\begin{aligned} z_1 &= b_{11}x_1 + \dots + b_{1n}x_n, \\ &\dots\dots\dots\dots \\ z_m &= b_{m1}x_1 + \dots + b_{mn}x_n. \end{aligned}$$

Сложим эти формы соответственно

$$\begin{aligned} y_1 + z_1 &= (a_{11} + b_{11})\,x_1 + \dots + (a_{1n} + b_{1n})\,x_n, \\ &\dots\dots\dots\dots\dots\dots\dots\dots \\ y_m + z_m &= (a_{m1} + b_{m1})\,x_1 + \dots + (a_{mn} + b_{mn})\,x_n. \end{aligned}$$

Естественно считать, что матрица получившейся системы форм

$$\begin{pmatrix} a_{11}+b_{11} & \dots & a_{1n}+b_{1n} \\ \dots & \dots & \dots \\ a_{m1}+b_{m1} & \dots & a_{mn}+b_{mn} \end{pmatrix}$$

является суммой матриц $\begin{pmatrix} a_{11} \dots a_{1n} \\ \dots \dots \\ a_{m1} \dots a_{mn} \end{pmatrix}$ и $\begin{pmatrix} b_{11} \dots b_{1n} \\ \dots \dots \\ b_{m1} \dots b_{mn} \end{pmatrix}$.

Аналогичным образом, произведение матрицы $\begin{pmatrix} a_{11} \dots a_{1n} \\ \dots \dots \\ a_{m1} \dots a_{mn} \end{pmatrix}$ на число c определяется как матрица коэффициентов в системе форм $cy_1, cy_2, \dots, cy_m$, где $y_1, y_2, \dots, y_m$ — формы, коэффициенты которых образуют матрицу $\begin{pmatrix} a_{11} \dots a_{1n} \\ \dots \dots \\ a_{m1} \dots a_{mn} \end{pmatrix}$.

Из этого определения видно, что

$$c\begin{pmatrix} a_{11} \dots a_{1n} \\ \dots \dots \\ a_{m1} \dots a_{mn} \end{pmatrix} = \begin{pmatrix} ca_{11} \dots ca_{1n} \\ \dots \dots \\ ca_{m1} \dots ca_{mn} \end{pmatrix}.$$

Наконец, действие умножения матрицы на матрицу определяется следующим образом. Пусть

$$\left.\begin{aligned} z_1 &= a_{11}y_1 + \dots + a_{1m}y_m, \\ &\dots\dots\dots\dots \\ z_k &= a_{k1}y_1 + \dots + a_{km}y_m \end{aligned}\right\} \qquad (1)$$

и

$$\begin{aligned} y_1 &= b_{11}x_1 + \dots + b_{1n}x_n, \\ &\dots\dots\dots\dots \\ y_m &= b_{m1}x_1 + \dots + b_{mn}x_n. \end{aligned}$$

Подставив в (1) выражения $y_1, y_2, \dots, y_m$ через $x_1, x_2, \dots, x_n$, получим, что $z_1, z_2, \dots, z_k$ выражаются через $x_1, x_2, \dots, x_n$ тоже в виде линейных форм

$$\begin{aligned} z_1 &= c_{11}x_1 + \dots + c_{1n}x_n, \\ &\dots\dots\dots\dots \\ z_k &= c_{k1}x_1 + \dots + c_{kn}x_n. \end{aligned}$$

Матрица коэффициентов $\begin{pmatrix} c_{11} \dots c_{1n} \\ \dots\dots \\ c_{k1} \dots c_{kn} \end{pmatrix}$ называется произведением матриц $\begin{pmatrix} a_{11} \dots a_{1m} \\ \dots\dots \\ a_{k1} \dots a_{km} \end{pmatrix}$ и $\begin{pmatrix} b_{11} \dots b_{1n} \\ \dots\dots \\ b_{m1} \dots b_{mn} \end{pmatrix}$ и обозначается через

$$\begin{pmatrix} a_{11} \dots a_{1m} \\ \dots\dots \\ a_{k1} \dots a_{km} \end{pmatrix} \begin{pmatrix} b_{11} \dots b_{1n} \\ \dots\dots \\ b_{m1} \dots b_{mn} \end{pmatrix}.$$

Нетрудно подсчитать, как выражаются элементы произведения двух матриц через элементы их сомножителей. Элемент c_{ij} есть коэффициент при x_j в выражении z_i через $x_1, \dots, x_n$.

Но $z_i = a_{i1}y_1 + \dots + a_{im}y_m$, а

$$y_1 = \dots + b_{1j}x_j + \dots,$$
$$\dots\dots\dots\dots$$
$$y_m = \dots + b_{mj}x_j + \dots$$

Следовательно,

$$z_i = \dots + (a_{i1}b_{1j} + \dots + a_{im}b_{mj})\,x_j + \dots,$$

откуда

$$c_{ij} = a_{i1}b_{1j} + \dots + a_{im}b_{mj}.$$

Итак, элемент i-й строки и j-го столбца произведения двух матриц равен сумме произведений элементов i-й строки первого сомножителя на соответствующие элементы j-го столбца второго сомножителя. Например:

$$\begin{pmatrix} 2 & 1 & 3 \\ 3 & 1 & 1 \end{pmatrix} \begin{pmatrix} 3 & 1 \\ 1 & 2 \\ 2 & 4 \end{pmatrix} = \begin{pmatrix} 2\cdot 3+1\cdot 1+3\cdot 2 & 2\cdot 1+1\cdot 2+3\cdot 4 \\ 3\cdot 3+1\cdot 1+1\cdot 2 & 3\cdot 1+1\cdot 2+1\cdot 4 \end{pmatrix} = \begin{pmatrix} 13 & 16 \\ 12 & 9 \end{pmatrix}.$$

Хотя матрица является, так сказать, «составным» объектом, в состав которого входит много элементов, полезно и удобно обозначать ее одной буквой, сохраняя при этом для действий сложения и умножения обычные обозначения. Мы будем пользоваться для обозначения матриц большими буквами латинского алфавита. Применение таких сокращенных обозначений вносит простоту и обозримость в теорию матриц, охватывая в кратких формулах, напоминающих формулы обычной алгебры, сложные соотношения, связывающие множество чисел — эле-

ментов матриц, участвующих в этих формулах. Так, например, совокупность линейных форм

$$\begin{aligned} &a_{11}x_1 + \ldots + a_{1n}x_n, \\ &\cdots\cdots\cdots\cdots \\ &a_{m1}x_1 + \ldots + a_{mn}x_n \end{aligned}$$

в матричных обозначениях выглядит как AX, где A — матрица коэффициентов, X — «столбец», составленный из переменных $x_1, \ldots, x_n$. Система линейных уравнений

$$\begin{aligned} &a_{11}x_1 + \ldots + a_{1n}x_n = b_1, \\ &\cdots\cdots\cdots\cdots \\ &a_{m1}x_1 + \ldots + a_{mn}x_n = b_m \end{aligned}$$

записывается как

$$AX = B,$$

где A — матрица коэффициентов, X — столбец из неизвестных, B — столбец из свободных членов.

Основные действия над матрицами — действия сложения и умножения — определены, конечно, не всегда. Действие сложения имеет смысл для матриц *одинакового строения*, т. е. имеющих одинаковое число строк и столбцов. В результате сложения получается матрица того же строения. Действие умножения имеет смысл, если число столбцов первой матрицы равно числу строк второй. В результате получается матрица, число строк которой равно числу строк первого сомножителя, а число столбцов равно числу столбцов второго сомножителя.

Действия над квадратными матрицами подчиняются большей части законов действий над числами, но некоторые законы оказываются нарушенными.

Перечислим основные свойства действий над матрицами:

1. $A + B = B + A$ (переместительный закон для сложения).
2. $(A + B) + C = A + (B + C)$ (сочетательный закон для сложения).
3. $c(A + B) = cA + cB$
3′. $(c_1 + c_2)A = c_1A + c_2A$

(распределительные законы при умножении на число. Здесь c, c_1, c_2 — числа, а не матрицы).

4. $(c_1c_2)A = c_1(c_2A)$ (сочетательный закон умножения на число).

5. Существует «нулевая» матрица $O = \begin{pmatrix} 0 \ldots 0 \\ \ldots \\ 0 \ldots 0 \end{pmatrix}$ такая, что $A + O = A$ при любой матрице A.

6. $c \cdot O = 0 \cdot A = O$; обратно, если $cA = O$, то $c = 0$ или $A = O$ (здесь c — число).

7. Для любой матрицы A существует противоположная матрица $-A$, т. е. такая, что $A+(-A)=O$.

8. $(A+B)\cdot C=AC+BC$
8′. $C(A+B)=CA+CB$ } (распределительные законы для сложения и умножения матриц).

9. $(AB)C=A(BC)$ (сочетательный закон для умножения).

10. $(cA)B=A(cB)=c(AB)$.

Эти свойства имеют место не только для квадратных, но и для любых прямоугольных матриц, с единственной оговоркой, что действия, входящие в каждую из перечисленных формул, должны быть определены. Для квадратных матриц одинакового порядка эта оговорка естественно отпадает.

Приведенные свойства действий аналогичны свойствам действий над числами.

Укажем теперь на две особенности действий над матрицами.

Во-первых, при умножении матриц, даже квадратных, переместительный закон может не иметь места, т. е. AB не всегда равно BA. Например:

$$\begin{pmatrix} 3 & -2 \\ -1 & 4 \end{pmatrix}\begin{pmatrix} 1 & 2 \\ 3 & 2 \end{pmatrix}=\begin{pmatrix} -3 & 2 \\ 11 & 6 \end{pmatrix};$$

$$\begin{pmatrix} 1 & 2 \\ 3 & 2 \end{pmatrix}\begin{pmatrix} 3 & -2 \\ -1 & 4 \end{pmatrix}=\begin{pmatrix} 1 & 6 \\ 7 & 2 \end{pmatrix}.$$

Во-вторых, известно, что произведение двух чисел равно нулю в том и только в том случае, если хотя бы один из сомножителей равен нулю. Эта теорема является, как известно, основной в теории алгебраических уравнений. При умножении матриц она оказывается неверной. Именно, произведение двух матриц может равняться нулевой матрице, хотя ни один из сомножителей не равен нулевой матрице. Например:

$$\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}\begin{pmatrix} 1 & 1 \\ -1 & -1 \end{pmatrix}=\begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}.$$

Отметим еще одно свойство умножения матриц. Матрица $\overline{A}$ будет называться *транспонированной* к матрице A, если в каждой строке $\overline{A}$ расположить элементы соответствующего столбца матрицы A с сохранением порядка. Например, для матрицы

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

транспонированной будет матрица

$$\overline{A}=\begin{pmatrix} 1 & 3 & 5 \\ 2 & 4 & 6 \end{pmatrix}.$$

Действие умножения матриц связано с операцией транспонирования формулой

$$\overline{AB} = \overline{B}\,\overline{A},$$

которая легко проверяется на основании правила умножения матриц.

Теория матриц составляет важную и неотъемлемую часть линейной алгебры, играя роль аппарата при постановке и решении ее задач.

Геометрические аналогии в линейной алгебре. Кроме описанного выше источника возникновения идей и задач линейной алгебры — потребностей математического анализа, геометрия, в частности аналитическая геометрия, тоже нуждается в развитии линейной алгебры и со своей стороны обогащает ее важными идеями и аналогиями. Известно, что аналитическая геометрия на плоскости и в еще большей мере в пространстве, в части, касающейся теории прямых и плоскостей, использует в простейшей форме аппарат линейной алгебры. Действительно, прямая линия на плоскости задается линейным уравнением с двумя переменными, которое связывает две координаты любой точки прямой. Плоскость в пространстве задается линейным уравнением с тремя переменными (координатами произвольной точки этой плоскости), прямая в пространстве — двумя линейными уравнениями.

Известно, что особенная простота и ясность вносятся в аналитическую геометрию, а следовательно, и в теорию простейших систем линейных уравнений, если использовать понятие вектора. Подобную же простоту и ясность вносит в линейную алгебру, в частности в общую теорию систем линейных уравнений, использование понятия вектора в некотором обобщенном смысле. Путь для этого обобщения следующий. Вектор (в пространстве) задается тремя числами — тремя проекциями на оси координат. Каждая тройка действительных чисел в свою очередь может быть изображена геометрически в виде вектора (в пространстве).

Для векторов определены действия сложения («по правилу параллелограма») и умножения на число. Эти действия определяются в соответствии с аналогичными действиями над силами, скоростями, ускорениями и другими физическими величинами, изображаемыми посредством векторов.

Если векторы задаются своими координатами (т. е. проекциями на координатные оси), то действиям сложения и умножения на число, производимым над векторами, соответствуют одноименные действия над строками (или столбцами) из их координат.

Таким образом, строку или столбец из трех элементов удобно геометрически интерпретировать как вектор в трехмерном пространстве, при этом основные действия над «строками» (или «столбцами») интерпретируются соответствующими действиями над векторами в пространстве, так что алгебра строк (или столбцов) из трех элементов формально

ничем не отличается от алгебры векторов трехмерного пространства. Это обстоятельство делает естественным введение в линейную алгебру геометрической терминологии.

Столбец (или строка) из n чисел $\begin{pmatrix} x_1 \\ x_2 \\ \cdot \\ \cdot \\ x_n \end{pmatrix}$ рассматривается как «вектор», т. е. как элемент некоторого «n-мерного векторного пространства».

Суммой векторов $\begin{pmatrix} x_1 \\ x_2 \\ \cdot \\ \cdot \\ x_n \end{pmatrix}$ и $\begin{pmatrix} y_1 \\ y_2 \\ \cdot \\ \cdot \\ y_n \end{pmatrix}$ считается вектор $\begin{pmatrix} x_1+y_1 \\ x_2+y_2 \\ \cdot\ \cdot \\ x_n+y_n \end{pmatrix}$; произведением вектора $\begin{pmatrix} x_1 \\ x_2 \\ \cdot \\ \cdot \\ x_n \end{pmatrix}$ на число c считается вектор $\begin{pmatrix} cx_1 \\ cx_2 \\ \cdot \\ \cdot \\ cx_n \end{pmatrix}$. Совокупность всех векторов (столбцов) составляет, по определению, n-мерное арифметическое векторное пространство.

Наряду с n-мерным арифметическим векторным пространством можно ввести понятие n-мерного точечного пространства, сопоставляя каждому столбцу из n действительных чисел геометрический образ — точку. Тогда n-мерное векторное пространство определяется следующим образом. Каждой паре точек A и B сопоставляем вектор $\overrightarrow{AB}$, идущий из точки A в точку B, считая, по определению, его координатами (проекциями на оси координат) разности соответствующих координат точек B и A. Два вектора считаются равными, если равны их соответствующие координаты, аналогично тому, как в трехмерном пространстве мы считаем векторы равными, если один из них получается из другого параллельным переносом.

Между векторами n-мерного векторного пространства и точками n-мерного точечного пространства естественным образом устанавливается взаимно однозначное соответствие.

Точка $\begin{pmatrix} 0 \\ 0 \\ \cdot \\ \cdot \\ 0 \end{pmatrix}$ принимается за «начало координат», и каждой точке соотносится вектор, соединяющий начало координат с этой точкой. Тогда каждому вектору сопоставляется точка, являющаяся концом этого вектора, в предположении, что его начало совмещено с началом координат. Введение точечного пространства создает новые аналогии, позволяющие лучше «видеть» в n-мерном пространстве.

Однако при дальнейших обобщениях (§ 2) строгое определение точечного пространства становится значительно более сложным и поэтому мы не будем использовать это понятие. Читателю, желающему пользоваться аналогиями, возникающими из рассмотрения точечного пространства, следует представлять себе элементы векторного пространства как векторы, исходящие из начала координат.

Введение геометрической терминологии дает возможность использовать в линейной алгебре аналогии, основанные на геометрической интуиции, создающейся при изучении геометрии трехмерного пространства.

Конечно, этими аналогиями нужно пользоваться с известной осторожностью, имея в виду возможность проверить строго логически каждое наглядно-геометрическое рассуждение и применяя только точные определения «геометрических» понятий и строго доказанные теоремы.

Характерной особенностью элементов n-мерного векторного пространства является наличие действий сложения и умножения на число, по своим свойствам напоминающих действия над числами. Именно, как уже было отмечено в описании свойств действий над матрицами, для действия сложения выполнены переместительный и сочетательный законы, верны распределительные законы (при умножении на число), действие сложения однозначно обратимо, произведение вектора на число дает нулевой вектор в том и только в том случае, если либо вектор есть нулевой вектор, либо число равно нулю.

Однако упомянутыми особенностями обладают не только столбцы (и строки). Ими обладают и совокупности матриц одинакового строения и физические векторные величины: силы, скорости, ускорения и т. д. Ими обладают и некоторые математические объекты совершенно другой природы, например: совокупность всех многочленов от одной переменной, совокупность всех непрерывных функций, заданных на данном отрезке $[a, b]$, совокупность всех решений линейного однородного дифференциального уравнения и т. д.

Указанное обстоятельство делает полезным дальнейшее обобщение векторного пространства, именно введение общих линейных пространств. Элементами таких обобщенных пространств могут быть любые математические или физические объекты, для которых некоторым естественным образом определены действия сложения и умножения на число. Такой весьма общий и отвлеченный подход к понятию линейного пространства не вносит, как мы увидим далее, никаких осложнений в теорию: любое линейное (конечно, n-мерное; что́ это значит, будет объяснено в следующем параграфе) пространство по своему строению и свойствам ничем не отличается от арифметического линейного пространства, но область приложений при этом обобщении значительно расширяется, появляется возможность применять методы линейной алгебры к весьма широкому кругу задач теоретического естествознания.

§ 2. ЛИНЕЙНОЕ ПРОСТРАНСТВО

Определение линейного пространства. Переходим к строгому определению линейного пространства.

Линейным пространством называется совокупность объектов любой природы, для которых имеют смысл понятия суммы и произведения на число, удовлетворяющих следующим требованиям:

1. $(X+Y)+Z=X+(Y+Z)$.
2. Существует «нулевой» элемент 0, такой, что $X+0=X$ при любом X.
3. Для любого элемента X существует противоположный $-X$, такой, что $X+(-X)=0$.
4. $X+Y=Y+X$.
5. $1\cdot X=X$.
6. $c_1(c_2X)=c_1c_2X$.
7. $(c_1+c_2)X=c_1X+c_2X$.
8. $c(X+Y)=cX+cY$.

Здесь X, Y, Z — элементы линейного пространства; 1, c_1, c_2, c — числа.

Эти требования (называемые иначе *аксиомами линейного пространства*) весьма естественны и представляют собой формальное описание тех свойств действий сложения и умножения на число, которые неотъемлемо связывают с понятиями этих действий, в каком бы обобщенном смысле они ни понимались. Действия, имеющие тот или иной физический смысл, трактуются как сложение и умножение на число именно в тех случаях, когда эти действия удовлетворяют требованиям 1—8.

Отметим некоторые следствия из этих аксиом.

а) Нулевой элемент 0 пространства единственен, т. е. существует только один элемент, удовлетворяющий аксиоме 2.

б) Для данного элемента X существует единственный противоположный элемент.

в) Имеет смысл «вычитание», т. е. по данной сумме и одному из слагаемых всегда определяется второе слагаемое и при этом только одно. Именно: если $X+Z=Y$, то $Z=Y+(-X)$.

г) $0\cdot X=c\cdot 0=0$.

д) Если $cX=0$, то или $c=0$, или $X=0$.

е) $-X=(-1)X$.

Доказательства этих следствий очень просты, и мы не будем на них останавливаться. В дальнейшем элементы линейного пространства мы будем называть векторами.

Линейная зависимость и независимость векторов. Перейдем теперь к важному понятию линейной зависимости и независимости векторов.

Линейной комбинацией векторов $X_1, X_2, \ldots, X_m$ называется вектор $c_1X_1+c_2X_2+\ldots+c_mX_m$ при некоторых численных значениях коэффициентов $c_1, c_2, \ldots, c_m$. Если среди векторов $X_1, X_2, \ldots, X_m$ найдется

хотя бы один, являющийся линейной комбинацией остальных, то векторы $X_1, X_2, \ldots, X_m$ называются линейно-зависимыми. Если же ни один из векторов $X_1, X_2, \ldots, X_m$ не является линейной комбинацией остальных, то векторы $X_1, X_2, \ldots, X_m$ называются линейно-независимыми.

Легко видеть, что для линейной независимости векторов $X_1, X_2, \ldots, X_m$ необходимо и достаточно, чтобы соотношение $c_1X_1 + c_2X_2 + \ldots + c_mX_m = 0$ выполнялось только при $c_1 = c_2 = \ldots = c_m = 0$.

Для векторов в обычном трехмерном пространстве понятия линейной зависимости и независимости имеют простой геометрический смысл.

Пусть имеются два вектора X_1 и X_2. Линейная зависимость их означает, что один из векторов является «линейной комбинацией» другого, т. е. просто отличается от него численным множителем. Это значит, что оба вектора укладываются на общей прямой, т. е. они имеют одинаковое или противоположное направление.

Обратно, если два вектора укладываются на одной прямой, то они линейно-зависимы. Следовательно, линейная независимость двух векторов X_1 и X_2 означает, что эти векторы не могут быть уложены на одну прямую; их направления существенно различны.

Рассмотрим теперь, что означает линейная зависимость и независимость трех векторов. Допустим, что векторы X_1, X_2 и X_3 линейно-зависимы, и положим для определенности, что вектор X_3 является линейной комбинацией векторов X_1 и X_2. Тогда X_3, очевидно, расположен в плоскости, содержащей векторы X_1 и X_2, т. е. все три вектора X_1, X_2, X_3 укладываются на одной плоскости. Легко видеть, что если векторы X_1, X_2, X_3 лежат в одной плоскости, то они линейно-зависимы. Действительно, если векторы X_1 и X_2 не лежат на одной прямой, то X_3 можно разложить по X_1 и X_2, т. е. представить X_3 в виде линейной комбинации X_1 и X_2. Если же X_1 и X_2 лежат на одной прямой, то уже X_1 и X_2 линейно-зависимы.

Итак, линейная зависимость трех векторов X_1, X_2, X_3 равносильна тому, что они лежат в одной плоскости. Следовательно, X_1, X_2, X_3 линейно-независимы в том и только в том случае, если они не укладываются на одной плоскости.

Четыре вектора в трехмерном пространстве всегда линейно-зависимы. Действительно, если векторы X_1, X_2, X_3 линейно-зависимы, то при любом X_4 векторы X_1, X_2, X_3, X_4 тоже линейно-зависимы. Если же X_1, X_2, X_3 линейно-независимы, то они не лежат в одной плоскости и любой вектор X_4 можно разложить по X_1, X_2, X_3, т. е. представить в виде их линейной комбинации.

Проведенные рассуждения можно обобщить следующим образом.

В трехмерном пространстве векторы $X_1, X_2, \ldots, X_k$ $(k \geqslant 3)$ линейно-зависимы в том и только в том случае, если они укладываются в пространство (прямая или плоскость) с числом измерений, меньшим k.

В дальнейшем, после строгого определения подпространства и размерности, мы увидим, что и в общем случае линейная зависимость векторов $X_1, X_2, \ldots, X_k$ равносильна тому, что они укладываются в пространство, размерность которого меньше k, т. е. «геометрический» смысл линейной зависимости остается тот же самый, что для векторов в трехмерном пространстве.

В теории линейных пространств играет существенную роль следующая теорема. Если векторы $X_1, X_2, \ldots, X_m$ являются линейными комбинациями векторов $Y_1, Y_2, \ldots, Y_k$ и $m > k$, то векторы $X_1, X_2, \ldots, X_m$ линейно-зависимы (теорема о линейной зависимости линейных комбинаций).

Для $k = 1$ теорема очевидна. Для $k > 1$ теорема легко доказывается методом математической индукции по числу k.

Базис и размерность пространства. В трехмерном пространстве любые три вектора X_1, X_2, X_3, не лежащие в одной плоскости (т. е. линейно-независимые), образуют *базис* этого пространства, это значит, что любой вектор пространства может быть разложен по векторам X_1, X_2, X_3, т. е. представлен в виде их линейной комбинации.

Общие линейные векторные пространства могут быть резко разделены на два типа.

Возможно, что в пространстве существует сколь угодно большое число линейно-независимых векторов. Такие пространства называются бесконечномерными и их изучение выходит за рамки линейной алгебры, являясь предметом специальной математической дисциплины — функционального анализа (см. главу XIX).

Линейное пространство называется конечномерным, если в нем существует конечная граница для числа линейно-независимых векторов, т. е. такое число n, что в пространстве существует n линейно-независимых векторов, но любые векторы в числе, большем n, линейно-зависимы. Число n называется *размерностью* или числом измерений пространства.

Так, пространство векторов обычного геометрического трехмерного пространства трехмерно и в смысле данного общего определения. Действительно, в трехмерном геометрическом пространстве существует сколько угодно троек линейно-независимых векторов, но любые четыре вектора уже линейно-зависимы.

Пространство n-членных столбцов n-мерно в смысле данного определения. Действительно, в этом пространстве существует n линейно-независимых векторов, например векторы

$$e_1 = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \quad e_2 = \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix}, \ldots, \quad e_n = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix}, \tag{2}$$

но всякий вектор $\begin{pmatrix} x_1 \\ x_2 \\ \cdot \\ \cdot \\ x_n \end{pmatrix}$ этого пространства является их линейной комбинацией, именно: $x_1e_1+x_2e_2+\ldots+x_ne_n$. Следовательно, в силу теоремы о линейной зависимости линейных комбинаций, любые векторы в числе, большем n, линейно-зависимы.

Многочлены от одной переменной образуют линейное пространство. В самом деле, для многочленов естественным образом определены действия сложения и умножения на число, удовлетворяющие аксиомам 1—8. Однако это пространство бесконечномерно, ибо векторы $1, x, \ldots, x^N$ линейно-независимы при любом N. Совокупность же многочленов, степени которых не превосходят данного числа N, образуют конечномерное пространство, размерность которого равна $N+1$. Действительно, векторы $1, x, \ldots, x^N$ линейно-независимы и число их равно $N+1$. Всякий же многочлен, степень которого не превосходит N, является линейной комбинацией $1, x, \ldots, x^N$, и, следовательно, по той же теореме о линейной зависимости любые многочлены степени $\leqslant N$, взятые в числе, большем $N+1$, линейно-зависимы.

Введем теперь важное понятие *базиса* для n-мерного пространства. Базисом называется такая совокупность линейно-независимых векторов пространства, что любой вектор пространства является линейной комбинацией векторов этой совокупности. Так, в пространстве столбцов базисом является, например, совокупность векторов (2). В пространстве многочленов степени $\leqslant N$ за базис можно принять «векторы» $1, x, \ldots, x^N$. В трехмерном геометрическом пространстве роль базиса играет любая тройка линейно-независимых векторов.

В n-мерном линейном пространстве любая совокупность из n линейно-независимых векторов (а существование хотя бы одной такой совокупности содержится в определении n-мерного пространства) образует базис пространства. Действительно, пусть $e_1, e_2, \ldots, e_n$ — линейно-независимые векторы n-мерного линейного пространства и X — какой-либо вектор пространства. Тогда векторы $X, e_1, \ldots, e_n$ линейно-зависимы (ибо их число больше n), т. е. найдутся числа $c, c_1, c_2, \ldots, c_n$, не равные нулю одновременно, такие, что $cX+c_1e_1+\ldots+c_ne_n=0$. При этом $c\neq 0$, ибо если бы $c=0$, векторы $e_1, e_2, \ldots, e_n$ были бы линейно-зависимыми. Следовательно, $X=-\frac{c_1}{c}e_1-\ldots-\frac{c_n}{c}e_n$, т. е. любой вектор пространства есть линейная комбинация векторов $e_1, e_2, \ldots, e_n$.

Любой базис n-мерного линейного пространства состоит ровно из n векторов. Действительно, векторы базиса линейно-независимы, и потому число их не может быть больше n. С другой стороны, пусть $e_1, e_2, \ldots, e_k$ какой-либо базис n-мерного пространства. Уже установлено,

что $k \leqslant n$. Но любой вектор пространства, по определению базиса, является линейной комбинацией векторов $e_1, e_2, \ldots, e_k$, и, в силу теоремы о линейной зависимости линейных комбинаций, любые векторы, взятые в числе, большем k, линейно-зависимы, откуда следует, что размерность пространства n не больше числа векторов базиса k. Итак, $k = n$, что и требовалось доказать.

Введем теперь *координаты* вектора относительно данного базиса $e_1, e_2, \ldots, e_n$. Как уже было сказано выше, любой вектор X является линейной комбинацией векторов базиса. Такое представление единственно. Действительно, пусть вектор X выражается через базис $e_1, e_2, \ldots, e_n$ двумя способами:

$$X = x_1 e_1 + x_2 e_2 + \ldots + x_n e_n,$$
$$X = x_1' e_1 + x_2' e_2 + \ldots + x_n' e_n.$$

Тогда $(x_1 - x_1') e_1 + (x_2 - x_2') e_2 + \ldots + (x_n - x_n') e_n = 0$, откуда следует, в силу линейной независимости $e_1, e_2, \ldots, e_n$, что $x = x_1', \ldots, x_n = x_n'$.

Коэффициенты $x_1, x_2, \ldots, x_n$ в разложении произвольного вектора X через векторы базиса называются *координатами* вектора X в этом базисе. Таким образом, любому вектору, *если только выбран базис пространства*, естественным образом сопоставляется строка (или столбец) из его координат, и наоборот: любая строка (или столбец) из n чисел может рассматриваться как совокупность координат некоторого вектора.

Действиям сложения векторов и умножения вектора на число соответствуют одноименные действия над строками (или столбцами) из координат.

Поэтому любое n-мерное линейное пространство, независимо от природы его элементов (будут они функциями, матрицами, какими-либо физическими величинами и т. д.), по отношению к этим действиям ничем не отличается от пространства строк (или столбцов). Таким образом, как уже было сказано раньше, обобщенный, аксиоматический подход к понятию линейного пространства не вносит никаких усложнений по сравнению с трактовкой пространства как пространства строк, но значительно расширяет круг приложений этого понятия.

Тождественность свойств двух совокупностей объектов по отношению к заданной системе действий (или каких-либо других отношений между элементами) называется в математике изоморфизмом. Точное определение изоморфизма алгебраических систем будет дано в главе XX. Пользуясь этим термином, мы можем сказать, что все n-мерные линейные пространства, независимо от природы их элементов, изоморфны друг другу и изоморфны единой модели — пространству строк.

Подпространства. Совокупность векторов n-мерного линейного пространства R_n, удовлетворяющая требованию, что каждая линейная ком-

бинация любых векторов рассматриваемой совокупности тоже принадлежит к ней, называется *подпространством* этого пространства. Очевидно, что подпространство пространства R_n само является линейным пространством и, следовательно, имеет базис и размерность. Очевидно также, что размерность подпространства не превосходит размерности всего пространства и может равняться ей в том и только в том случае, когда подпространство совпадает со всем пространством.

Примерами подпространств трехмерного векторного пространства могут служить рассматриваемые с точностью до переноса плоскости и прямые, точнее, совокупности всех векторов, укладывающихся на одной плоскости или на одной прямой.

Наиболее часто приходится рассматривать подпространства, «натянутые» на систему векторов. Эти подпространства определяются следующим образом. Пусть дана система линейно-независимых или зависимых векторов $X_1, X_2, \ldots, X_m$ пространства R_n. Тогда совокупность всех линейных комбинаций этих векторов $\{c_1X_1 + c_2X_2 + \ldots + c_mX_m\}$ образует подпространство пространства R_n, которое и называется подпространством, натянутым на векторы $X_1, X_2, \ldots, X_m$.

Размерность такого подпространства называется *рангом* системы векторов $X_1, X_2, \ldots, X_m$. Легко видеть, что ранг системы векторов равен максимальному числу линейно-независимых векторов, содержащихся в этой системе.

«Совокупность», состоящая только из нулевого вектора, формально удовлетворяет требованиям, предъявляемым к подпространству. Размерность этого подпространства считается равной нулю.

Если даны два подпространства пространства R_n, то из них естественным образом конструируются еще два подпространства — их векторная сумма и пересечение.

Векторной суммой двух подпространств P и Q называется совокупность всех сумм векторов, принадлежащих подпространствам P и Q. Векторную сумму можно рассматривать также как подпространство, натянутое на объединение базисов подпространств P и Q.

Пересечением двух подпространств называется совокупность всех векторов, принадлежащих обоим подпространствам. Например, векторной суммой двух плоскостей (т. е. двумерных векторных подпространств) в обычном трехмерном пространстве является все пространство (если только плоскости не совпадают), а пересечением — прямая линия (при той же оговорке).

Размерности p и q двух данных подпространств, размерность t их векторной суммы и размерность s их пересечения удовлетворяют следующему интересному соотношению:

$$p + q = t + s.$$

Доказательство этого утверждения мы опускаем.

Из этого соотношения можно сделать некоторые выводы, касающиеся пересечения подпространств в частных случаях. Например, две несовпадающие плоскости (т. е. двумерные подпространства) в пространстве четырех измерений пересекаются, вообще говоря, только в точке (размерность их пересечения равна нулю) и две плоскости пересекаются по прямой только в том случае, если их векторная сумма трехмерна, т. е. если обе плоскости укладываются в некоторое трехмерное подпространство. Действительно, в этом случае $t+s=2+2=4$, откуда следует, что $s=1$ только при $t=3$.

Комплексное линейное пространство. При описании пространства строк и общего линейного пространства мы не уточняли, о каких числах идет речь, когда определяется действие умножения вектора на число. Так как мы исходили из обобщения обычных векторов, т. е. направленных отрезков в геометрическом трехмерном пространстве, то мы имели в виду любые действительные числа. Построенные таким образом линейные пространства, называемые действительными линейными пространствами, наиболее естественным образом обобщают трехмерное пространство обычных векторов. Однако для многих задач современной математики оказывается полезным рассмотрение *комплексного линейного пространства*. Под этим термином понимается совокупность объектов, для которых определены действия сложения и умножения на любое комплексное число, причем эти действия удовлетворяют всем аксиомам 1—8. Примером комплексного пространства может служить пространство строк, элементами которых являются любые комплексные числа.

Формально теория комплексных пространств ничем существенно не отличается от теории действительных пространств.

Однако даже двумерное комплексное пространство не имеет наглядной геометрической интерпретации. Дело в том, что n-мерное комплексное пространство можно рассматривать и как действительное, ибо поскольку для его элементов определено действие умножения на любые комплексные числа, тем самым определено и действие умножения на действительное число. Но размерность комплексного n-мерного пространства, рассматриваемого как действительное, будет равна $2n$, т. е. вдвое больше. В самом деле, если $\boldsymbol{e}_1, \boldsymbol{e}_2, \ldots, \boldsymbol{e}_n$ есть базис комплексного пространства, то за базис этого пространства, рассматриваемого как действительное, можно принять, например, векторы

$$\boldsymbol{e}_1, i\boldsymbol{e}_1, \boldsymbol{e}_2, i\boldsymbol{e}_2, \ldots, \boldsymbol{e}_n, i\boldsymbol{e}_n, \quad \text{где } i=\sqrt{-1}.$$

Следовательно, двумерное комплексное пространство может быть интерпретировано как действительное, но четырехмерное.

Далее, теория линейных пространств не претерпевает формально никаких изменений, если в качестве совокупности чисел, на которые

допускается умножение «векторов» пространства, брать какие-либо другие совокупности чисел, отличные от совокупности всех действительных и всех комплексных чисел, лишь бы результаты основных арифметических действий — сложения, вычитания, умножения и деления — над числами этой совокупности снова к ней принадлежали. Совокупности чисел, удовлетворяющие этим требованиям, носят название числовых полей. (Более подробно это понятие рассматривается в главе XX.) Примером числового поля может служить, например, поле рациональных чисел.

В некоторых разделах алгебры, близких к теории чисел, с успехом применяется теория линейных пространств над произвольным полем.

n-мерное эвклидово пространство. В предыдущем изложении еще не были обобщены некоторые важные понятия обычного векторного пространства, в частности понятия длины вектора и угла между векторами. Известно, что в аналитической геометрии в вопросах, касающихся пересечения прямых и плоскостей, параллельности и многих других, эти понятия и не используются. Свойства пространства, описание которых не нуждается в понятиях длины вектора и угла, могут быть охарактеризованы как свойства, не нарушающиеся при любых аффинных преобразованиях [см. главу III (том 1), § 11]. По этой причине линейные пространства, в которых не определены понятия длины вектора, называются *аффинными пространствами*.

Однако многие задачи математики требуют обобщения понятий длины вектора и угла на n-мерные пространства. Эти обобщения производятся посредством аналогий с теорией векторов на плоскости и в пространстве.

Рассмотрим сначала действительное пространство строк. Длина вектора $X=(x_1, x_2, \ldots, x_n)$ считается, по определению, равной числу $|X| = \sqrt{x_1^2+x_2^2+\ldots+x_n^2}$. Это совершенно естественно, так как при $n=2$ и $n=3$ именно по этой формуле вычисляется длина вектора через его координаты относительно декартовых осей координат.

Понятие угла между векторами вводится естественным способом из следующих соображений. На плоскости и в пространстве угол между векторами X и Y есть угол при вершине A в треугольнике со сторонами $AB=|X|$, $AC=|Y|$ и $BC=|X-Y|$.

Для n-мерного пространства естественно принять это за определение угла между векторами, т. е. как бы считать, что пару векторов можно «извлечь» из n-мерного пространства и «положить» на плоскость с сохранением их длин и угла между ними. Однако в таком определении имеется нестрогость: существование треугольника ABC с длинами векторов $|X|$, $|Y|$ и $|X-Y|$ нуждается в доказательстве.

Пренебрегая этой неточностью, выведем формулу для вычисления угла. По известной формуле тригонометрии

$$BC^2 = AB^2 + AC^2 - 2AB \cdot AC \cos\varphi,$$

откуда

$$\cos\varphi=\frac{|X|^2+|Y|^2-|X-Y|^2}{2|X|\cdot|Y|}=$$

$$=\frac{x_1^2+\ldots+x_n^2+y_1^2+\ldots+y_n^2-(x_1-y_1)^2-\ldots-(x_n-y_n)^2}{2|X|\cdot|Y|}=$$

$$=\frac{x_1y_1+\ldots+x_ny_n}{|X|\cdot|Y|}.$$

Если, как и для трехмерного пространства, для произведения длин векторов на косинус угла между ними сохранить термин «скалярное произведение», то мы получим, что скалярное произведение векторов вычисляется по формуле

$$X\cdot Y=x_1y_1+\ldots+x_ny_n,$$

которая при $n=2$ и $n=3$ совпадает с известной формулой для скалярного произведения обычных векторов.

Строго говоря, выражение $x_1y_1+\ldots+x_ny_n$ нужно принять за определение *скалярного произведения* (ибо в определении скалярного произведения при помощи угла была допущена нестрогость), а затем определить угол между векторами по формуле

$$\cos\varphi=\frac{X\cdot Y}{|X|\cdot|Y|}. \tag{3}$$

Так мы и сделаем.

Для законности такого определения угла необходимо доказать, что правая часть формулы (3) по абсолютной величине не превосходит 1, т. е. что $(X\cdot Y)^2\leqslant|X|^2\cdot|Y|^2$.

В развернутой форме это неравенство имеет вид

$$(x_1y_1+\ldots+x_ny_n)^2\leqslant(x_1^2+\ldots+x_n^2)(y_1^2+\ldots+y_n^2).$$

Оно носит название неравенства Коши—Буняковского и может быть доказано непосредственным, но довольно громоздким вычислением. Мы докажем его путем следующего косвенного рассуждения.

Прежде всего отметим, что скалярное умножение векторов обладает следующими свойствами:

1′. $X\cdot X=|X|^2>0$ при $X\neq0$.

2′. $X\cdot Y=Y\cdot X$.

3′. $(cX)\cdot Y=c(X\cdot Y)$.

4′. $(X_1+X_2)\cdot Y=X_1\cdot Y+X_2\cdot Y$.

Выполнение этих свойств непосредственно следует из выражения скалярного произведения через координаты.

Введем теперь в рассмотрение вектор $Y+tX$, где t — любое действительное число. Имеем $|Y+tX|^2\geqslant0$, ибо квадрат длины вектора не может быть отрицательным. Но в силу свойств скалярного произ-

ведения $|Y+tX|^2=|Y|^2+2tX\cdot Y+t^2|X|^2$. Далее известно, что квадратный трехчлен может оставаться неотрицательным при всех значениях действительной переменной t только в том случае, если его корни мнимые или равные, т. е. если его дискриминант отрицателен или равен нулю. Но дискриминант трехчлена $|Y|^2+2tX\cdot Y+t^2|X|^2$ равен $4(X\cdot Y)^2-4|X|^2|Y|^2$, и, следовательно, $(X\cdot Y)^2-|X|^2|Y|^2\leqslant 0$, что равносильно неравенству Коши—Буняковского.

Из доказанного неравенства вытекает, что $\frac{|X\cdot Y|}{|X||Y|}\leqslant 1$, и, следовательно, определение угла по формуле (3) законно.

Далее, легко выводятся неравенства

$$|X|-|Y|\leqslant|X\pm Y|\leqslant|X|+|Y|,$$

из которых, в частности, следует существование треугольника со сторонами $|X|$, $|Y|$ и $|X-Y|$, так что данное выше нестрогое, но геометрически наглядное определение угла также приобретает законную силу.

Аксиоматическое определение n-мерного эвклидова пространства. В предыдущем пункте мы ввели понятие длины вектора, угла и скалярного произведения в пространстве строк. В общем аксиоматически определенном n-мерном действительном линейном пространстве эти понятия определяются тоже аксиоматически, причем в основу кладется понятие скалярного произведения.

Скалярным умножением векторов линейного действительного пространства называется сопоставление каждой паре векторов X и Y действительного числа, называемого их скалярным произведением $X\cdot Y$, причем это сопоставление должно удовлетворять следующим требованиям (аксиомам):

1′. $X\cdot X>0$ при $X\neq 0$, $0\cdot 0=0$.

2′. $X\cdot Y=Y\cdot X$.

3′. $(cX)\cdot Y=c(X\cdot Y)$.

4′. $(X_1+X_2)\cdot Y=X_1\cdot Y+X_2\cdot Y$.

Далее, за длину вектора принимается число $\sqrt{X\cdot X}$, за косинус угла между векторами X и Y — число, равное $\frac{X\cdot Y}{|X|\cdot|Y|}$. Для законности этого последнего определения необходимо установить справедливость неравенства Коши—Буняковского $(X\cdot Y)^2\leqslant|X|^2|Y|^2$. Но это делается совершенно так же, как было сделано в предыдущем пункте. В проведенном доказательстве как раз и были использованы только свойства 1′, 2′, 3′ и 4′ скалярного произведения, специфика пространства строк в этом доказательстве не играла никакой роли. Действительное линейное пространство, в котором введено скалярное умножение, удовлетворяющее аксиомам 1′—4′, называется *эвклидовым пространством*.

В различных конкретных линейных пространствах, изучаемых в математике, скалярное произведение вводится различными способами, выбор которых обусловливается существом вопроса. Например, в пространствах, элементами которых служат функции от одной переменной $X(t)$, определенные на заданном промежутке $a \leqslant t \leqslant b$, за скалярное произведение двух элементов $X(t)$ и $Y(t)$ часто принимается число $\int_a^b X(t)Y(t)\,dt$ или число $\int_a^b X(t)Y(t)p(t)\,dt$, где $p(t)$ — некоторая положительная функция. Легко видеть, что при каждом из этих определений все аксиомы 1′—4′ выполняются.

Ортогональность. Ортонормальный базис. Два вектора эвклидова пространства называются *ортогональными* (или перпендикулярными), если их скалярное произведение равно нулю. Легко видеть, что попарно-ортогональные ненулевые векторы всегда линейно-независимы. Действительно, пусть $X_1, X_2, \ldots, X_m$ — попарно-ортогональные ненулевые векторы, и пусть $c_1X_1 + c_2X_2 + \ldots + c_mX_m = 0$.

По свойству скалярного произведения $X_1(c_1X_1 + c_2X_2 + \ldots + c_mX_m) = c_1|X_1|^2 = 0$, откуда $c_1 = 0$. Таким же образом доказывается, что $c_2 = \ldots = c_m = 0$. Следовательно, $X_1, \ldots, X_m$ линейно-независимы.

Из доказанного следует, что в n-мерном пространстве может существовать не более n попарно-ортогональных ненулевых векторов, и каждая совокупность из n попарно-ортогональных векторов образует базис пространства. Если, кроме того, длины всех n попарно-ортогональных векторов равны единице, то образованный ими базис называется *ортонормальным*.

Нетрудно доказать, но мы на этом не будем останавливаться, что в эвклидовом пространстве существуют ортонормальные базисы и их даже бесконечно много. Более того, если в пространстве R выбрано некоторое подпространство P, то ортонормальный базис подпространства можно дополнить посредством присоединения нескольких векторов до ортонормального базиса всего пространства.

Векторы в эвклидовом пространстве удобнее всего задавать координатами в каком-либо ортонормальном базисе, так как в этом случае получается особенно простое выражение для скалярного произведения. Действительно, если вектор X имеет координаты $(x_1, x_2, \ldots, x_n)$ в ортонормальном базисе $e_1, e_2, \ldots, e_n$, а вектор Y — координаты $(y_1, y_2, \ldots, y_n)$, т. е.

$$X = x_1e_1 + x_2e_2 + \ldots + x_ne_n \text{ и } Y = y_1e_1 + y_2e_2 + \ldots + y_ne_n,$$

то по свойству скалярного произведения

$$\begin{aligned} X \cdot Y = {} & x_1y_1e_1e_1 + x_1y_2e_1e_2 + \ldots + x_1y_ne_1e_n + \\ & + x_2y_1e_2e_1 + x_2y_2e_2e_2 + \ldots + x_2y_ne_2e_n + \\ & \ldots\ldots\ldots\ldots\ldots\ldots \\ & + x_ny_1e_ne_1 + x_ny_2e_ne_2 + \ldots + x_ny_ne_ne_n = \\ = {} & x_1y_1 + x_2y_2 + \ldots + x_ny_n, \end{aligned}$$

так как $e_i e_k = 0$ при $i \neq k$ и $e_i e_i = |e_i|^2 = 1$ при любом $i = 1, 2, \ldots, n$. В частности, $X \cdot X = x_1^2 + x_2^2 + \ldots + x_n^2$.

Таким образом, длина вектора и скалярное произведение выражаются через координаты ортонормального базиса по тем же формулам, что и в пространстве строк.

Переход от одной из моделей эвклидова пространства — пространства строк — к общему аксиоматически определенному эвклидову пространству не влечет за собой никаких осложнений, но только расширяет область приложения теории.

Рассмотрим еще вопрос об ортогональном проектировании векторов на подпространство. Пусть R_n — некоторое n-мерное эвклидово пространство и P_m — его m-мерное подпространство. Пусть, далее, $e_1, e_2, \ldots, e_m, f_1, \ldots, f_{n-m}$ — ортонормальный базис R_n, включающий ортонормальный базис подпространства P_m. Подпространство Q_{n-m}, натянутое на векторы $f_1, f_2, \ldots, f_{n-m}$, называется *ортогонально-дополнительным* к подпространству P_m. Его размерность равна $n - m$. Ортогонально-дополнительное подпространство Q_{n-m} может быть охарактеризовано как совокупность всех векторов, ортогональных ко всем векторам подпространства P_m.

Любой вектор Z, принадлежащий R_n, может быть однозначно представлен в виде суммы векторов X и Y, из которых один принадлежит подпространству P_m, другой — подпространству Q_{n-m}. Это ясно из возможности и однозначности представления вектора Z в виде

$$Z = x_1 e_1 + \ldots + x_m e_m + y_1 f_1 + \ldots + y_{n-m} f_{n-m},$$

так что $X = x_1 e_1 + \ldots + x_m e_m$, $Y = y_1 f_1 + \ldots + y_{n-m} f_{n-m}$.

Вектор X называется *ортогональной проекцией вектора* Z на P_m.

Унитарное пространство. Понятия длины вектора и скалярного произведения векторов могут быть определены и в комплексном пространстве. В основу попрежнему кладется понятие скалярного умножения, которое определяется следующим образом. Каждой паре X и Y векторов комплексного пространства сопоставляется комплексное (не обязательно действительное) число, называемое их скалярным произведением $X \cdot Y$. Действие скалярного умножения должно удовлетворять следующим аксиомам:

1″. $X \cdot X$ — действительное и положительное при $X \neq 0$, $0 \cdot 0 = 0$.

2″. $Y \cdot X = (X \cdot Y)'$.

Здесь штрих означает переход к комплексно-сопряженному числу.

3″. $(cX) \cdot Y = c(X \cdot Y)$ при любом комплексном c.

4″. $(X_1 + X_2) \cdot Y = X_1 \cdot Y + X_2 \cdot Y$ (распределительный закон).

В пространстве строк с комплексными элементами за скалярное произведение векторов $(x_1, \ldots, x_n)$ и $(y_1, \ldots, y_n)$ можно принять число $x_1 y_1' + \ldots + x_n y_n'$. Легко проверить, что при таком определении все аксиомы 1″—4″ выполнены.

За длину вектора принимают число $\sqrt{X \cdot X}$. Понятие угла между векторами не определяется.

Комплексное линейное пространство со скалярным произведением, удовлетворяющим аксиомам 1″—4″, называется *унитарным пространством*.

§ 3. СИСТЕМЫ ЛИНЕЙНЫХ УРАВНЕНИЙ

Системы двух уравнений с двумя неизвестными и трех уравнений с тремя неизвестными. Система двух линейных уравнений с двумя неизвестными выглядит в общем виде так:

$$a_1x + b_1y = c_1,$$
$$a_2x + b_2y = c_2.$$

Для решения этой системы умножим первое уравнение на b_2, второе на $-b_1$ и сложим. Получим

$$(a_1b_2 - a_2b_1)\,x = c_1b_2 - c_2b_1.$$

Аналогичным образом, умножив первое уравнение на $-a_2$, второе на a_1 и сложив, получим

$$(a_1b_2 - a_2b_1)\,y = a_1c_2 - a_2c_1.$$

Из этих равенств легко определяются x и y, если только выражение $a_1b_2 - a_2b_1$, оказывающееся коэффициентом при неизвестных x и y, отлично от нуля. Это выражение называется *определителем* матрицы $\begin{pmatrix} a_1 & b_1 \\ a_2 & b_2 \end{pmatrix}$, образованной коэффициентами системы. Определитель обозначается так: $\begin{vmatrix} a_1 & b_1 \\ a_2 & b_2 \end{vmatrix}$. Из данного определения следует, что определитель вычисляется по схеме

не требующей дальнейших пояснений.

Вернемся к решению системы. Выражения $c_1b_2 - c_2b_1$ и $a_1c_2 - a_2c_1$, согласно данному определению, тоже представляют собой определители, именно: $\begin{vmatrix} c_1 & b_1 \\ c_2 & b_2 \end{vmatrix}$ и $\begin{vmatrix} a_1 & c_1 \\ a_2 & c_2 \end{vmatrix}$.

Таким образом, если определитель $\begin{vmatrix} a_1 & b_1 \\ a_2 & b_2 \end{vmatrix}$ отличен от нуля, мы получаем следующие формулы для решения системы:

$$x = \frac{\begin{vmatrix} c_1 & b_1 \\ c_2 & b_2 \end{vmatrix}}{\begin{vmatrix} a_1 & b_1 \\ a_2 & b_2 \end{vmatrix}}; \quad y = \frac{\begin{vmatrix} a_1 & c_1 \\ a_2 & c_2 \end{vmatrix}}{\begin{vmatrix} a_1 & b_1 \\ a_2 & b_2 \end{vmatrix}}. \tag{4}$$

Строго говоря, проведенные рассуждения не полны. Действия над уравнениями, которые мы производили при выводе формул для решения системы, были осмыслены только в предположении, что x и y уже представляют собой числа, образующие решение системы. Логическая сущность проведенного рассуждения такова: если определитель из коэффициентов системы не равен нулю и решение системы существует, то оно вычисляется по формулам (4). Поэтому еще необходимо установить, что найденные значения для неизвестных действительно удовлетворяют обоим уравнениям системы. Это делается без труда.

Итак, если определитель матрицы из коэффициентов системы отличен от нуля, то система имеет единственное решение, которое дается формулами (4).

Для системы трех уравнений с тремя неизвестными

$$\begin{aligned} a_1x + b_1y + c_1z &= d_1, \\ a_2x + b_2y + c_2z &= d_2, \\ a_3x + b_3y + c_3z &= d_3 \end{aligned}$$

нетрудно провести аналогичные рассуждения и выкладки — для этого достаточно сложить уравнения, умножив их предварительно на такие множители, чтобы после сложения уничтожались сразу два неизвестных. В качестве множителей для уничтожения неизвестных y и z следует взять $b_2c_3 - b_3c_2$, $b_3c_1 - b_1c_3$ и $b_1c_2 - b_2c_1$, что легко проверить вычислением.

В результате получится, что если выражение

$$\Delta = a_1b_2c_3 - a_1b_3c_2 + a_2b_3c_1 - a_2b_1c_3 + a_3b_1c_2 - a_3b_2c_1$$

отлично от нуля, то система имеет единственное решение, получаемое по формулам

$$x = \frac{\Delta_1}{\Delta}, \quad y = \frac{\Delta_2}{\Delta}, \quad z = \frac{\Delta_3}{\Delta},$$

где Δ_1, Δ_2, Δ_3 — выражения, получающиеся из Δ посредством замены коэффициентов при соответствующих неизвестных свободными членами.

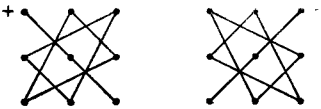
Выражение Δ называется определителем матрицы

$$\begin{pmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{pmatrix}$$

и обозначается через

$$\begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix}.$$

Для вычисления определителя удобна следующая схема:



На первой из этих схем соединены линиями (диагональ и два треугольника) места, где находятся элементы, произведения которых входят в состав определителя со знаком плюс; на второй схеме — то же для слагаемых, входящих в состав определителя со знаком минус.

Мы получили для систем двух уравнений с двумя неизвестными и трех уравнений с тремя неизвестными совершенно сходные результаты. В обоих случаях система имеет единственное решение, если определитель матрицы коэффициентов отличен от нуля. Формулы для решений тоже аналогичны: в знаменателе каждой из неизвестных находится определитель матрицы коэффициентов, а в числителях — определители матриц, получающихся из матриц коэффициентов заменой коэффициентов при вычисляемой неизвестной свободными членами.

Непосредственное обобщение этих результатов на системы n уравнений с n неизвестными при любом n затруднительно. Это сравнительно легко удается сделать косвенными средствами: сперва обобщить понятие определителя на квадратные матрицы любого порядка и, изучив свойства определителей, применить их теорию к исследованию систем.

Определитель n-го порядка. Рассматривая развернутое выражение для определителей

$$\begin{vmatrix} a_1 & b_1 \\ a_2 & b_2 \end{vmatrix} = a_1 b_2 - a_2 b_1$$

и

$$\begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix} = a_1 b_2 c_3 - a_1 b_3 c_2 + a_2 b_3 c_1 - a_2 b_1 c_3 + a_3 b_1 c_2 - a_3 b_2 c_1,$$

замечаем, что в каждое слагаемое входят в качестве сомножителей по одному элементу из каждой строки и по одному из каждого столбца определителя, причем всевозможные произведения этого вида входят в состав определителя со знаком плюс или минус. Это свойство полагается в основу обобщения понятия определителя на квадратные матрицы любого порядка. Именно: определителем квадратной матрицы n-го порядка или, короче, *определителем n-го порядка* называется алгебраическая сумма всевозможных произведений элементов матрицы, взятых по одному из каждой строки и по одному из каждого столбца, причем полученные произведения снабжены знаками плюс и минус по некоторому вполне определенному правилу. Это правило вводится

довольно сложным образом, и мы не будем останавливаться на его формулировке. Существенно отметить, что оно устанавливается так, что обеспечивается следующее важнейшее основное свойство определителя:

1. При перестановке двух строк определитель меняет знак на противоположный.

Для определителя 2 и 3-го порядков это свойство легко проверяется непосредственным вычислением. В общем случае оно доказывается на основе не сформулированного нами здесь правила знаков.

Определители обладают целым рядом других замечательных свойств, которые дают возможность с успехом использовать определители в разнообразных теоретических и численных расчетах, несмотря на чрезвычайную громоздкость определителя: ведь определитель n-го порядка содержит, как нетрудно видеть, $n!$ слагаемых, каждое слагаемое состоит из n сомножителей и слагаемые снабжены знаками по некоторому сложному правилу.

Переходим к перечислению основных свойств определителей, не останавливаясь на их подробных доказательствах.

Первое из этих свойств уже сформулировано выше.

2. Определитель не меняется при транспонировании его матрицы, т. е. при замене строк на столбцы с сохранением порядка.

Доказательство основано на подробном исследовании правила расстановки знаков в слагаемых определителя. Это свойство дает возможность всякое утверждение, касающееся строк определителя, перенести на столбцы.

3. Определитель есть линейная функция от элементов какой-либо его строки (или столбца). Подробнее

$$\begin{vmatrix} a_{11} & \dots & a_{1n} \\ \dots & \dots & \dots \\ a_{i1} & \dots & a_{in} \\ \dots & \dots & \dots \\ a_{n1} & \dots & a_{nn} \end{vmatrix} = a_{i1}A_{i1} + a_{i2}A_{i2} + \dots + a_{in}A_{in}, \tag{5}$$

где $A_{i1}, A_{i2}, \dots, A_{in}$ представляют собой выражения, не зависящие от элементов i-й строки.

Это свойство с очевидностью следует из того, что каждое слагаемое содержит по одному и только одному сомножителю из каждой, в частности i-й, строки.

Равенство (5) называется разложением определителя по элементам i-й строки, а коэффициенты $A_{i1}, A_{i2}, \dots, A_{in}$ называются алгебраическими дополнениями элементов $a_{i1}, a_{i2}, \dots, a_{in}$ в определителе.

4. Алгебраическое дополнение A_{ij} элемента a_{ij} равно, с точностью до знака, так называемому минору Δ_{ij} определителя, т. е. определи-

телю $(n-1)$-го порядка, получающемуся из данного посредством вычеркивания i-й строки и j-го столбца. Для получения алгебраического дополнения минор нужно взять со знаком $(-1)^{i+j}$. Свойства 3 и 4 сводят вычисление определителя порядка n к вычислению n определителей порядка $n-1$.

Из перечисленных основных свойств вытекает ряд интересных свойств определителей. Перечислим некоторые из них.

5. Определитель с двумя одинаковыми строками равен нулю.

Действительно, если определитель имеет две одинаковые строки, то при их перестановке определитель не изменяется, ибо строки одинаковые, но вместе с тем он, в силу первого свойства, меняет знак на обратный. Следовательно, он равен нулю.

6. Сумма произведений элементов какой-либо строки на алгебраические дополнения другой строки равна нулю.

Действительно, такая сумма является результатом разложения определителя с двумя одинаковыми строками по одной из них.

7. Общий множитель элементов какой-либо строки можно вынести за знак определителя.

Это следует из свойства 3.

8. Определитель с двумя пропорциональными строками равен нулю.

Достаточно вынести множитель пропорциональности, и мы получим определитель с двумя равными строками.

9. Определитель не меняется, если к элементам какой-либо строки добавить числа, пропорциональные элементам другой строки.

Действительно, в силу свойства 3 преобразованный определитель равен сумме исходного определителя и определителя с двумя пропорциональными строками, который равен нулю.

Последнее свойство дает хорошее средство для вычисления определителей. Используя это свойство можно, не меняя величины определителя, преобразовать его матрицу так, чтобы в какой-либо строке (или столбце) все элементы, кроме одного, оказались равными нулю. Затем, разложив определитель по элементам этой строки (столбца), мы сведем вычисление определителя порядка n к вычислению *одного* определителя порядка $n-1$, именно, алгебраического дополнения единственного отличного от нуля элемента выбранной строки.

Например, нужно вычислить определитель

$$\Delta = \begin{vmatrix} 1 & 1 & -1 & 2 \\ 2 & -1 & 1 & 1 \\ -1 & 2 & 0 & 1 \\ 1 & 1 & -2 & 1 \end{vmatrix}$$

Добавив ко второму столбцу первый, умноженный на —1, к третьему — первый, к четвертому — первый, умноженный на —2, получим

$$\Delta = \begin{vmatrix} 1 & 0 & 0 & 0 \\ 2 & -3 & 3 & -3 \\ -1 & 3 & -1 & 3 \\ 1 & 0 & -1 & -1 \end{vmatrix}$$

Разложив Δ по элементам первой строки, получим

$$\Delta = 1 \cdot (-1)^{1+1} \begin{vmatrix} -3 & 3 & -3 \\ 3 & -1 & 3 \\ 0 & -1 & -1 \end{vmatrix}$$

Наконец, добавив к первой строке вторую и разложив по элементам первого столбца, получим

$$\Delta = \begin{vmatrix} 0 & 2 & 0 \\ 3 & -1 & 3 \\ 0 & -1 & -1 \end{vmatrix} = 3 \cdot (-1)^{1+2} \begin{vmatrix} 2 & 0 \\ -1 & -1 \end{vmatrix} = -3 \cdot (-2) = 6.$$

Определитель матрицы A обозначают через $|A|$.

В заключение отметим ещё одно очень важное свойство определителей.

Определитель произведения двух квадратных матриц равен произведению определителей сомножителей, т. е. в сокращенной записи $|AB| = |A|\,|B|$.

Это свойство дает, в частности, возможность умножать определители одинакового порядка по правилу умножения матриц.

Системы n линейных уравнений с n неизвестными. Теперь, используя аппарат определителей, легко обобщить результаты, полученные ранее для систем двух уравнений с двумя неизвестными и трех уравнений с тремя неизвестными, на системы n уравнений с n неизвестными в предположении, что определитель матрицы коэффициентов отличен от нуля.

Пусть

$$\begin{gathered} a_{11}x_1 + \ldots + a_{1j}x_j + \ldots + a_{1n}x_n = b_1, \\ \ldots\ldots\ldots\ldots\ldots\ldots \\ a_{n1}x_1 + \ldots + a_{nj}x_j + \ldots + a_{nn}x_n = b_n \end{gathered}$$

такая система. Обозначим через Δ определитель матрицы коэффициентов системы. Он, по предположению, отличен от нуля. Далее, через A_{ij} обозначим алгебраическое дополнение элемента a_{ij}. Умножим первое уравнение на A_{1j}, второе — на A_{2j}, ..., n-е — на A_{nj} и сложим. Получим

$$\Delta x_j = b_1 A_{1j} + \ldots + b_n A_{nj}.$$

Действительно, коэффициенты при всех неизвестных, кроме x_j, обращаются в 0, ибо они представляют собой суммы произведений алгебраи-

ческих дополнений элементов j-го столбца на элементы других столбцов (свойство 6, примененное к столбцам); коэффициент же при неизвестном x_j равен сумме произведений элементов j-го столбца на их алгебраические дополнения, т. е. равен Δ.

Таким образом,

$$x_j = \frac{b_1 A_{1j} + \ldots + b_n A_{nj}}{\Delta} \qquad \text{при всех } j = 1, 2, \ldots, n. \tag{6}$$

Как уже говорилось выше, приведенные рассуждения осмыслены только в случае, если под $x_1, x_2, \ldots, x_n$ подразумевать решение системы, существование которого тем самым необходимо предполагать.

Таким образом, результатом рассуждения явилось следующее.

Если решение системы существует, то оно дается формулами (6) и тем самым единственно.

Для полноты изложения необходимо доказать существование решения, что достигается подстановкой найденных значений для неизвестных во все уравнения исходной системы. Легко убедиться, используя то же свойство определителя (в применении к строкам), что найденные значения действительно удовлетворяют всем уравнениям.

Итак, верна следующая теорема: если определитель матрицы коэффициентов системы n уравнений с n неизвестными отличен от нуля, то система имеет единственное решение, даваемое формулами (6).

Эти формулы можно еще преобразовать, заметив, что сумму $b_1 A_{1j} + \ldots + b_n A_{nj}$ можно записать в виде определителя, именно:

$$\Delta_j = b_1 A_{1j} + \ldots + b_n A_{nj} = \begin{vmatrix} a_{11} & \ldots & b_1 & \ldots & a_{1n} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ a_{n1} & \ldots & b_n & \ldots & a_{nn} \end{vmatrix}$$

(свободные члены находятся в j-м столбце).

Таким образом, результаты, изложенные выше для систем уравнений с двумя и тремя неизвестными, полностью обобщены на систему n уравнений и даже формулы для решения получаются совершенно такими же по форме.

Отметим одно следствие из доказанной теоремы:

Если относительно системы уравнений заведомо известно, что она совсем не имеет решения или решение не единственно, то определитель матрицы из коэффициентов равен нулю.

Это следствие особенно часто применяется к однородным системам, т. е. таким, в которых свободные члены $b_1, b_2, \ldots, b_n$ все равны нулю. Однородные системы всегда имеют само собой разумеющееся «тривиальное» решение $x_1 = x_2 = \ldots = x_n = 0$.

Если же однородная система имеет, кроме тривиального, еще нетривиальное решение, то ее определитель равен нулю.

Это утверждение открывает возможность использовать теорию определителей в других разделах математики и в ее приложениях.

Рассмотрим, например, задачу из аналитической геометрии.

Найти уравнение плоскости, проходящей через три данные точки (x_1, y_1, z_1), (x_2, y_2, z_2) и (x_3, y_3, z_3), не лежащие на одной прямой.

Из элементарной геометрии известно, что искомая плоскость существует. Пусть ее уравнение имеет вид $Ax + By + Cz + D = 0$. Тогда

$$\begin{aligned} Ax_1 + By_1 + Cz_1 + D &= 0, \\ Ax_2 + By_2 + Cz_2 + D &= 0, \\ Ax_3 + By_3 + Cz_3 + D &= 0. \end{aligned}$$

Пусть x, y, z — координаты любой точки, лежащей на плоскости. Тогда и

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

Рассмотрим эти четыре уравнения как систему линейных однородных уравнений относительно коэффициентов A, B, C, D искомой плоскости. Эта система имеет нетривиальное решение, ибо искомая плоскость существует. Следовательно, определитель системы равен нулю, т. е.

$$\begin{vmatrix} x_1 & y_1 & z_1 & 1 \\ x_2 & y_2 & z_2 & 1 \\ x_3 & y_3 & z_3 & 1 \\ x & y & z & 1 \end{vmatrix} = 0. \tag{7}$$

Это и есть уравнение искомой плоскости. Действительно, оно есть уравнение 1-й степени относительно x, y, z, что следует из линейности определителя относительно элементов последней строки.

Используя то, что данные точки не лежат на одной прямой, нетрудно проверить, что не все его коэффициенты равны нулю. Следовательно, равенство (7) есть действительно уравнение плоскости. Плоскость эта проходит через данные точки, ибо координаты их, очевидно, удовлетворяют ее уравнению.

Матричная запись системы n уравнений с n неизвестными. Система n линейных уравнений с n неизвестными

$$\begin{aligned} a_{11}x_1 + \ldots + a_{1n}x_n &= b_1, \\ \ldots\ldots\ldots\ldots\ldots & \\ a_{n1}x_1 + \ldots + a_{nn}x_n &= b_n \end{aligned}$$

может быть записана в матричных обозначениях в виде одного равенства

$$AX = B.$$

Здесь A обозначает матрицу из коэффициентов, X — столбец, составленный из неизвестных, B — столбец из свободных членов.

Решение системы (если определитель матрицы A отличен от нуля) подробно записывается так [см. формулу (6)]:

$$x_1 = \frac{A_{11}}{\Delta} b_1 + \frac{A_{21}}{\Delta} b_2 + \ldots + \frac{A_{n1}}{\Delta} b_n,$$
$$x_2 = \frac{A_{12}}{\Delta} b_1 + \frac{A_{22}}{\Delta} b_2 + \ldots + \frac{A_{n2}}{\Delta} b_n,$$
$$\ldots\ldots\ldots\ldots\ldots\ldots$$
$$x_n = \frac{A_{1n}}{\Delta} b_1 + \frac{A_{2n}}{\Delta} b_2 + \ldots + \frac{A_{nn}}{\Delta} b_n$$

или в матричной форме

$$X = \begin{pmatrix} \frac{A_{11}}{\Delta} & \frac{A_{21}}{\Delta} & \ldots & \frac{A_{n1}}{\Delta} \\ \frac{A_{12}}{\Delta} & \frac{A_{22}}{\Delta} & \ldots & \frac{A_{n2}}{\Delta} \\ \ldots & \ldots & \ldots & \ldots \\ \frac{A_{1n}}{\Delta} & \frac{A_{2n}}{\Delta} & \ldots & \frac{A_{nn}}{\Delta} \end{pmatrix} B.$$

Матрица, стоящая первым множителем в правой части равенства, называется матрицей, *обратной* к матрице A, и обозначается через A^{-1}. Применяя это обозначение, получим решение системы $AX = B$ в следующем простом и естественном виде, напоминающем формулу для решения одного уравнения с одним неизвестным:

$$X = A^{-1}B.$$

Нетрудно дать другое обоснование полученному результату в терминах алгебры матриц.

Для этого прежде всего следует отметить особую роль матрицы

$$E = \begin{pmatrix} 1 & 0 & 0 & \ldots & 0 \\ 0 & 1 & 0 & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & 0 & \ldots & 1 \end{pmatrix},$$

называемой *единичной матрицей*.

Единичная матрица среди квадратных матриц играет такую же роль, какую играет число 1 среди всех чисел. Именно: при любой матрице A имеют место равенства $AE = A$ и $EA = A$. Это легко проверяется на основании правила умножения матриц.

Определенная выше матрица A^{-1}, обратная к матрице A, играет по отношению к ней роль, сходную с той, которую играет число, обратное к данному числу. Именно:

$$AA^{-1} = A^{-1}A = E.$$

Справедливость этих равенств проверяется на основании правил умножения матриц и свойств 3 и 6 определителя.

Зная эти свойства единичной и обратной матриц, можно провести решение системы $AX=B$ следующим образом.

Пусть $AX=B$. Тогда $A^{-1}(AX)=A^{-1}B$. Но $A^{-1}(AX)=(A^{-1}A)X=EX=X$, и, следовательно, $X=A^{-1}B$.

Пусть теперь $X=A^{-1}B$. Тогда $AX=AA^{-1}B=EB=B$.

Итак, «уравнение» $AX=B$ имеет единственное решение $X=A^{-1}B$, если только A^{-1} существует.

Мы установили существование обратной матрицы A^{-1} для матрицы A в предположении, что определитель матрицы A отличен от нуля. Это условие не только достаточно, но и необходимо для существования обратной матрицы. Действительно, пусть для матрицы A существует обратная A^{-1}, т. е. такая, что $AA^{-1}=E$. Тогда по свойству определителя произведения двух матриц

$$|A|\,|A^{-1}|=|E|=1,$$

откуда следует, что определитель матрицы A не равен нулю.

Матрица, определитель которой отличен от нуля, называется *невырожденной* или *неособенной*. Мы установили, таким образом, что обратная матрица всегда существует для невырожденных матриц и только для них.

Введение понятия обратной матрицы оказывается полезным не только в теории систем линейных уравнений, но и во многих других задачах линейной алгебры.

В заключение отметим, что выведенные формулы для решения линейных систем являются незаменимым орудием в теоретических исследованиях, но мало применимы для численного решения систем.

Как мы уже отмечали, для численного решения систем разработано много различных способов и вычислительных схем, и ввиду большой важности этой задачи для практики исследовательская работа по упрощению численного решения систем (особенно с большим числом неизвестных) интенсивно ведется и в настоящее время.

Общий случай систем линейных уравнений. Обратимся к исследованию систем линейных уравнений в самом общем случае, даже не предполагая, что число уравнений равно числу неизвестных. В такой общей постановке нельзя, естественно, ожидать, что решение системы всегда существует или в случае существования оно окажется единственным. Естественно предполагать, что если число уравнений меньше числа неизвестных, то система имеет бесконечно много решений. Например, двум уравнениям 1-й степени с тремя неизвестными удовлетворяют координаты любой из точек прямой линии, являющейся линией пересечения плоскостей, определяемых уравнениями. Однако может быть, что в этом случае система совсем не имеет решений, именно когда плоскости параллельны. Если же число уравнений больше числа неизвестных, то система, как правило, решений не имеет. Однако и в этом

случае возможно, что система имеет решения и даже бесконечно много.

Для исследования вопроса о существовании и характере множественности решений системы в такой общей постановке обратимся к «геометрическому» истолкованию системы.

Систему уравнений

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n &= b_1, \\ a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n &= b_2, \\ \ldots\ldots\ldots\ldots\ldots\ldots & \\ a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n &= b_m \end{aligned} \tag{8}$$

интерпретируем в m-мерном пространстве столбцов в виде

$$x_1\boldsymbol{A}_1 + x_2\boldsymbol{A}_2 + \ldots + x_n\boldsymbol{A}_n = \boldsymbol{B}.$$

Здесь $\boldsymbol{A}_1, \boldsymbol{A}_2, \ldots, \boldsymbol{A}_n$ обозначают столбцы из коэффициентов при соответствующих неизвестных, $\boldsymbol{B}$ — столбец из свободных членов.

В этой интерпретации вопрос о существовании решения системы превращается в вопрос о том, является ли данный вектор $\boldsymbol{B}$ линейной комбинацией векторов $\boldsymbol{A}_1, \boldsymbol{A}_2, \ldots, \boldsymbol{A}_n$.

Ответ на этот вопрос почти очевиден. Чтобы вектор $\boldsymbol{B}$ был линейной комбинацией векторов $\boldsymbol{A}_1, \boldsymbol{A}_2, \ldots, \boldsymbol{A}_n$, необходимо и достаточно, чтобы вектор $\boldsymbol{B}$ содержался в подпространстве, натянутом на $\boldsymbol{A}_1, \boldsymbol{A}_2, \ldots, \boldsymbol{A}_n$, или, иными словами, чтобы подпространства, натянутые на системы векторов $\boldsymbol{A}_1, \boldsymbol{A}_2, \ldots, \boldsymbol{A}_n$ и $\boldsymbol{A}_1, \boldsymbol{A}_2, \ldots, \boldsymbol{A}_n, \boldsymbol{B}$ совпадали.

Так как первое из этих подпространств содержится во втором, то для их совпадения необходимо и достаточно, чтобы были равны их размерности. Напомним, что размерность подпространства, натянутого на данную систему векторов, называется рангом этой системы векторов. Таким образом, необходимым и достаточным условием существования решения системы $x_1\boldsymbol{A}_1 + x_2\boldsymbol{A}_2 + \ldots + x_n\boldsymbol{A}_n = \boldsymbol{B}$ является равенство рангов систем векторов $\boldsymbol{A}_1, \boldsymbol{A}_2, \ldots, \boldsymbol{A}_n$ и $\boldsymbol{A}_1, \boldsymbol{A}_2, \ldots, \boldsymbol{A}_n, \boldsymbol{B}$.

Можно доказать, но мы на этом не будем останавливаться, что ранг системы векторов равен рангу матрицы, составленной из координат этих векторов. Здесь под *рангом матрицы* понимается наивысший из порядков отличных от нуля определителей, которые можно составить из данной матрицы посредством вычеркивания части ее строк и столбцов.

Так как координатами векторов $\boldsymbol{A}_1, \boldsymbol{A}_2, \ldots, \boldsymbol{A}_n$ (в естественном для пространства столбцов базисе) являются коэффициенты системы, а координатами вектора $\boldsymbol{B}$ — ее свободные члены, то мы получаем следующую окончательную формулировку условия существования решения системы.

Для существования по крайней мере одного решения системы линейных уравнений

$$a_{11}x_1 + \ldots + a_{1n}x_n = b_1,$$
$$\cdots\cdots\cdots$$
$$a_{m1}x_1 + \ldots + a_{mn}x_n = b_m$$

необходимо и достаточно, чтобы ранг матрицы, составленной из коэффициентов системы, равнялся рангу матрицы, составленной из коэффициентов и свободных членов.

Теперь исследуем характер множественности решений, если они существуют. Пусть $x_1^0, x_2^0, \ldots, x_n^0$ — какое-либо решение системы (8). Положим $x_1 = x_1^0 + y_1$, $x_2 = x_2^0 + y_2, \ldots, x_n = x_n^0 + y_n$. Тогда в силу того, что $x_1^0, x_2^0, \ldots, x_n^0$ образует решение системы (8), новые неизвестные $y_1, y_2, \ldots, y_n$ должны удовлетворять однородной системе

$$\begin{aligned} a_{11}y_1 + \ldots + a_{1n}y_n &= 0, \\ \cdots\cdots\cdots & \\ a_{m1}y_1 + \ldots + a_{mn}y_n &= 0 \end{aligned} \qquad (9)$$

с той же матрицей коэффициентов. Обратно, если к исходному решению $x_1^0, x_2^0, \ldots, x_n^0$ системы (8) добавить любое решение однородной системы (9), то получится снова решение системы (8).

Таким образом, чтобы получить общее решение системы (8), нужно взять какое-либо ее частное решение и сложить с общим решением однородной системы (9).

Тем самым вопрос о характере множественности решений системы (8) сводится к тому же вопросу для однородной системы (9). Этот вопрос мы рассмотрим в следующем пункте.

Однородные системы. Однородную систему линейных уравнений

$$\begin{aligned} a_{11}y_1 + a_{12}y_2 + \ldots + a_{1n}y_n &= 0, \\ a_{21}y_1 + a_{22}y_2 + \ldots + a_{2n}y_n &= 0, \\ \cdots\cdots\cdots\cdots & \\ a_{m1}y_1 + a_{m2}y_2 + \ldots + a_{mn}y_n &= 0 \end{aligned}$$

мы будем интерпретировать в n-мерном эвклидовом пространстве. (Тем самым мы предполагаем, что коэффициенты системы действительные. Для систем с комплексными коэффициентами можно дать аналогичную интерпретацию в унитарном пространстве, причем получатся аналогичные результаты).

Пусть $A_1', A_2', \ldots A_m'$, Y — векторы эвклидова пространства, координатами которых в некотором ортонормальном базисе являются соответственно

$$(a_{11}, a_{12}, \ldots, a_{1n}),\ (a_{21}, a_{22}, \ldots, a_{2n}),\ \ldots,\ (a_{m1}, a_{m2}, \ldots, a_{mn}),$$
$$(y_1, y_2, \ldots, y_n).$$

Тогда система принимает вид

$$A_1'Y=0,\quad A_2'Y=0,\ \ldots,\ A_m'Y=0,$$

т. е. каждое решение системы определяет вектор, ортогональный ко всем векторам, интерпретирующим коэффициенты отдельных уравнений.

Следовательно, множество решений образует подпространство, ортогонально-дополнительное к подпространству, натянутому на векторы A_1', A_2', ..., A_m'. Размерность последнего подпространства равна рангу r матрицы, составленной из коэффициентов системы. Размерность же ортогонального дополнения, т. е. «пространства решений», равна $n-r$.

Во всяком подпространстве существует базис, т. е. система линейно-независимых векторов, в числе, равном размерности подпространства, линейными комбинациями которых заполняется все подпространство. Следовательно, среди решений однородной системы существует $n-r$ линейно-независимых решений, таких, что все решения системы являются их линейными комбинациями. Здесь n обозначает число неизвестных, r — ранг матрицы из коэффициентов.

Таким образом, строение решений однородной системы, а следовательно, и неоднородной, выяснено до конца. В частности, однородная система имеет единственное тривиальное решение $x_1=x_2=\ldots=x_n=0$ в том и только в том случае, если ранг матрицы из коэффициентов равен числу неизвестных. В силу сказанного в конце предыдущего пункта, то же условие (при выполнении условия совместности) является условием единственности решения и для системы неоднородных уравнений.

Исследования систем, проведенные нами, наглядно показывают, как введение обобщенных геометрических понятий вносит простоту и обозримость в сложный алгебраический вопрос.

§ 4. ЛИНЕЙНЫЕ ПРЕОБРАЗОВАНИЯ

Определение и примеры. Во многих математических исследованиях возникает потребность в замене переменных, т. е. в переходе от одной системы переменных $x_1, x_2, \ldots, x_n$ к другой $y_1, y_2, \ldots, y_n$, связанной с первой посредством функциональной зависимости:

$$\begin{aligned} y_1&=\varphi_1(x_1, x_2, \ldots, x_n),\\ y_2&=\varphi_2(x_1, x_2, \ldots, x_n),\\ &\cdots\cdots\cdots\\ y_n&=\varphi_n(x_1, x_2, \ldots, x_n). \end{aligned}$$

Например, если переменные представляют собой координаты точки на плоскости или в пространстве, то переход от одной системы координат к другой системе влечет за собой преобразование координат, которые задаются выражениями исходных координат через новые, или наоборот.

Кроме того, преобразование переменных возникает при изучении изменений, переходов от одного положения или состояния к другому для таких объектов, положение или состояние которых описывается значениями переменных. Типичным примером этого рода преобразования может служить изменение координат точек некоторого тела при его деформации.

Отвлеченно заданное преобразование системы n переменных обычно интерпретируется именно как преобразование (деформация) n-мерного пространства, т. е. как сопоставление каждому вектору пространства (или его части) с координатами $x_1, x_2, \ldots, x_n$ соответствующего ему вектора с координатами $y_1, y_2, \ldots, y_n$.

Как уже было сказано выше, каждая «гладкая» (имеющая непрерывные частные производные) функция от нескольких переменных при малых изменениях этих переменных близка к линейной функции. Поэтому любое «гладкое» преобразование (т. е. такое, в аналитическом описании которого функции $\varphi_1, \varphi_2, \ldots, \varphi_n$ имеют непрерывные частные производные) на малой части пространства близко к линейному:

$$\begin{aligned} y_1 &= a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n + b_1, \\ y_2 &= a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n + b_2, \\ &\cdots\cdots\cdots\cdots\cdots\cdots \\ y_n &= a_{n1}x_1 + a_{n2}x_2 + \ldots + a_{nn}x_n + b_n. \end{aligned} \tag{10}$$

Уже одно это обстоятельство делает изучение свойств линейных преобразований одной из важнейших задач математики. Например, из теории n линейных уравнений с n неизвестными мы знаем, что необходимым и достаточным условием однозначной разрешимости системы уравнений (10) относительно $x_1, x_2, \ldots, x_n$, т. е. обратимости соответствующего линейного преобразования, является неравенство нулю определителя из коэффициентов. Это обстоятельство лежит в основе глубокой теоремы анализа: для того чтобы преобразование

$$\begin{aligned} y_1 &= \varphi_1(x_1, x_2, \ldots, x_n), \\ y_2 &= \varphi_2(x_1, x_2, \ldots, x_n), \\ &\cdots\cdots\cdots\cdots \\ y_n &= \varphi_n(x_1, x_2, \ldots, x_n), \end{aligned}$$

гладкое в окрестности данной точки, имело гладкое обратное преобразование, необходимо и достаточно, чтобы в данной точке определитель

$$\begin{vmatrix} \frac{\partial \varphi_1}{\partial x_1} & \cdots & \frac{\partial \varphi_1}{\partial x_n} \\ \frac{\partial \varphi_2}{\partial x_1} & \cdots & \frac{\partial \varphi_2}{\partial x_n} \\ \cdots & \cdots & \cdots \\ \frac{\partial \varphi_n}{\partial x_1} & \cdots & \frac{\partial \varphi_n}{\partial x_n} \end{vmatrix}$$

был отличен от нуля.

Изучение общего линейного преобразования (10) в основном сводится к изучению однородного преобразования с теми же коэффициентами

$$\begin{aligned} y_1 &= a_{11}x_1 + \ldots + a_{1n}x_n, \\ y_2 &= a_{21}x_1 + \ldots + a_{2n}x_n, \\ &\cdots\cdots\cdots\cdots \\ y_n &= a_{n1}x_1 + \ldots + a_{nn}x_n, \end{aligned} \tag{11}$$

и в дальнейшем, говоря о линейных преобразованиях, мы всегда будем подразумевать однородные преобразования.

Линейные преобразования n-мерного пространства могут быть также определены своими внутренними свойствами, помимо формул (11), связывающих координаты соответствующих точек. Такое бескоординатное определение понятия линейного преобразования полезно тем, что оно не зависит от выбора базиса. Это определение состоит в следующем.

Линейное преобразование n-мерного линейного пространства есть функция $\boldsymbol{Y} = \boldsymbol{A}(\boldsymbol{X})$, аргумент $\boldsymbol{X}$ и значения $\boldsymbol{Y}$ которой являются векторами. Функция эта удовлетворяет требованию линейности

$$\boldsymbol{A}(c_1\boldsymbol{X}_1 + c_2\boldsymbol{X}_2) = c_1\boldsymbol{A}(\boldsymbol{X}_1) + c_2\boldsymbol{A}(\boldsymbol{X}_2). \tag{12}$$

В дальнейшем, говоря о линейном преобразовании пространства, мы будем его понимать именно в смысле этого определения.

Это определение равносильно предыдущему — координатному. Действительно, функция $\boldsymbol{Y} = \boldsymbol{A}(\boldsymbol{X})$, сопоставляющая вектор $\boldsymbol{X}$ с координатами $x_1, x_2, \ldots, x_n$ вектор $\boldsymbol{Y}$ с координатами $y_1, y_2, \ldots, y_n$ так, что при этом координаты $y_1, y_2, \ldots, y_n$ выражаются через координаты $x_1, x_2, \ldots, x_n$ в виде линейных однородных функций, очевидно, удовлетворяет требованию (12). Обратно, если функция $\boldsymbol{Y} = \boldsymbol{A}(\boldsymbol{X})$ удовлетворяет требованию (12) и $\boldsymbol{e}_1, \boldsymbol{e}_2, \ldots, \boldsymbol{e}_n$ — какой-либо базис пространства, то

$$\boldsymbol{A}(x_1\boldsymbol{e}_1 + x_2\boldsymbol{e}_2 + \ldots + x_n\boldsymbol{e}_n) = x_1\boldsymbol{A}(\boldsymbol{e}_1) + x_2\boldsymbol{A}(\boldsymbol{e}_2) + \ldots + x_n\boldsymbol{A}(\boldsymbol{e}_n).$$

Обозначим координаты (в том же базисе) вектора $\boldsymbol{A}(\boldsymbol{e}_j)$ через $a_{1j}, \ldots, a_{nj}$; $j = 1, \ldots, n$. Тогда координаты вектора $\boldsymbol{Y} = \boldsymbol{A}(\boldsymbol{X})$ будут

$$\begin{aligned} y_1 &= a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n, \\ y_2 &= a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n, \\ &\cdots\cdots\cdots\cdots \\ y_n &= a_{n1}x_1 + a_{n2}x_2 + \ldots + a_{nn}x_n. \end{aligned}$$

Таким образом, каждому линейному преобразованию линейного пространства соответствует относительно данного базиса некоторая квадратная матрица. Само преобразование записывается на языке матриц в виде $Y = AX$. Здесь X — столбец из координат исходного вектора, Y — столбец из координат преобразованного вектора, A — матрица коэффициентов преобразования. Столбцы матрицы A образованы координатами тех векторов, в которые преобразуются векторы базиса. В соответствии с матричной записью мы будем в дальнейшем и само линейное преобразование часто записывать в виде $Y = AX$, опуская скобки.

Из формулы

$$A(x_1e_1 + x_2e_2 + \ldots + x_ne_n) = x_1A(e_1) + x_2A(e_2) + \ldots + x_nA(e_n)$$

следует, что все пространство при линейном преобразовании превращается в подпространство, натянутое на векторы $A(e_1), \ldots, A(e_n)$. Размерность этого подпространства равна рангу системы векторов $A(e_1)$, $A(e_2), \ldots, A(e_n)$, или, что то же самое, рангу матрицы, составленной из их координат, т. е. рангу матрицы A, сопоставляемой преобразованию. Подпространство это совпадает со всем пространством в том и только в том случае, когда ранг матрицы A равен n, т. е. когда определитель матрицы A отличен от нуля. В этом случае линейное преобразование называется *неособенным* или *невырожденным*.

Из теории систем линейных уравнений мы знаем, что невырожденные преобразования однозначно обратимы, и координаты исходного вектора выражаются через координаты преобразованного по формуле $X = A^{-1}Y$.

Преобразование, матрица которого имеет равный нулю определитель, называется *особенным* или *вырожденным*. Вырожденное преобразование необратимо. Это следует из теории линейных уравнений или более наглядно из того, что оно преобразует все пространство в его часть.

Примером невырожденного преобразования может служить в первую очередь единичное преобразование, преобразующее все векторы в себя. Матрицей единичного преобразования в любом базисе является единичная матрица E.

Невырожденными являются также преобразования подобия, состоящие в том, что все векторы пространства умножаются на одно и то же число. Матрица преобразования подобия не зависит от выбора базиса и имеет вид aE, где a — коэффициент подобия.

Важным частным случаем невырожденных преобразований являются преобразования ортогональные. Понятие *ортогонального преобразования* имеет смысл в применении к эвклидову пространству и определяется как линейное преобразование, сохраняющее длины векторов. Ортогональное преобразование есть обобщение на n-мерное пространство преобразования вращения пространства при неподвижном начале

координат или вращения, соединенного с отражением относительно какой-либо плоскости, проходящей через начало.

Легко видеть, что при ортогональном преобразовании сохраняются не только длины векторов, но и скалярные произведения, и, следовательно, ортогональные преобразования переводят ортонормальный базис пространства в систему попарно-ортогональных единичных векторов, которая в свою очередь неминуемо является базисом.

Матрица, связанная с ортогональным преобразованием, относительно ортонормального базиса обладает следующими специфическими свойствами.

Во-первых, сумма квадратов элементов каждого столбца равна 1, ибо такие суммы суть квадраты длин векторов, в которые переходят векторы выбранного базиса. Во-вторых, суммы произведений соответствующих элементов, взятых из двух различных столбцов, равны нулю, ибо такие суммы суть скалярные произведения векторов, в которые переходят векторы базиса.

В матричных обозначениях оба эти свойства записываются одной формулой

$$\bar{P}P = E.$$

Здесь P — матрица ортогонального преобразования (относительно ортонормального базиса), $\bar{P}$ — транспонированная с ней матрица, т. е. такая, строки которой являются столбцами матрицы P с сохранением их порядка.

В самом деле, диагональные элементы матрицы $\bar{P}P$ по правилу умножения матриц равны сумме квадратов элементов соответствующего столбца матрицы P, а недиагональные равны сумме произведений соответствующих элементов, взятых из различных столбцов матрицы P.

Примером вырожденного преобразования может служить, например, ортогональное проектирование всех векторов эвклидова пространства на некоторое подпространство (см. § 2). Действительно, при этом преобразовании все пространство отображается на свою часть.

Преобразование координат. Рассмотрим теперь вопрос о преобразовании координат в n-мерном пространстве, т. е. вопрос о том, как изменяются координаты векторов при переходе от одного базиса к другому.

Пусть дан исходный базис $e_1, e_2, \ldots, e_n$ и пусть $f_1, f_2, \ldots, f_n$ — какой-либо другой базис пространства. Пусть далее $C = \begin{pmatrix} c_{11}, \ldots, c_{1n} \\ \ldots\ldots \\ c_{n1}, \ldots, c_{nn} \end{pmatrix}$ — матрица, столбцы которой являются координатами векторов нового базиса $f_1, f_2, \ldots, f_n$ относительно исходного. Матрица C, очевидно, невырожденная в силу линейной независимости векторов $f_1, f_2, \ldots, f_n$. Она называется *матрицей преобразования координат*.

Обозначим через $x_1, x_2, \ldots, x_n$ координаты некоторого вектора X относительно базиса $e_1, e_2, \ldots, e_n$ и через $x'_1, x'_2, \ldots, x'_n$ — координаты того же вектора относительно базиса $f_1, f_2, \ldots, f_n$. Тогда $X = x'_1 f_1 + x'_2 f_2 + \ldots + x'_n f_n$, и, следовательно, координаты вектора X относительно исходного базиса образуют столбец

$$\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} c_{11}x'_1 + c_{12}x'_2 + \ldots + c_{1n}x'_n \\ c_{21}x'_1 + c_{22}x'_2 + \ldots + c_{2n}x'_n \\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot \\ c_{n1}x'_1 + c_{n2}x'_2 + \ldots + c_{nn}x'_n \end{pmatrix} = \begin{pmatrix} c_{11} \ldots c_{1n} \\ c_{21} \ldots c_{2n} \\ \cdot\ \cdot\ \cdot\ \cdot \\ c_{n1} \ldots c_{nn} \end{pmatrix} \begin{pmatrix} x'_1 \\ x'_2 \\ \vdots \\ x'_n \end{pmatrix}.$$

Итак, исходные координаты выражаются через преобразованные линейным однородным образом с матрицей C.

Формулы, выражающие зависимость между координатами относительно исходного и преобразованного базисов, формально совпадают с формулами, связывающими координаты соответствующих векторов при невырожденном линейном преобразовании пространства. Это обстоятельство дает возможность интерпретировать отвлеченно заданное линейное однородное преобразование переменных с невырожденной матрицей или как преобразование координат, или как линейное преобразование пространства. В каждом конкретном случае выбор одной из этих двух интерпретаций определяется содержанием рассматриваемой задачи.

Рассмотрим теперь вопрос о том, как изменяется матрица линейного преобразования пространства при преобразовании координат.

Пусть в базисе $e_1, e_2, \ldots, e_n$ данное линейное преобразование имеет матрицу A, так что столбец Y из координат преобразованного вектора связан со столбцом X исходного формулой

$$Y = AX.$$

Пусть теперь сделано преобразование координат с матрицей C; X', Y' обозначают соответственно столбцы из координат исходного и преобразованного векторов относительно нового базиса. Тогда $X = CX'$, $Y = CY'$, откуда

$$Y' = C^{-1}Y = C^{-1}AX = C^{-1}ACX'.$$

Итак, матрицей рассматриваемого преобразования относительно нового базиса является матрица $C^{-1}AC$.

Матрицы A и B, связанные соотношением $B = C^{-1}AC$, где C — некоторая неособенная матрица, называются *подобными*. Одному и тому же линейному преобразованию по отношению к различным базисам соответствует класс попарно подобных между собой матриц.

Собственные векторы и собственные значения линейного преобразования. Важнейший класс линейных преобразований образуют преобразования, осуществляемые следующим образом.

Пусть $e_1, e_2, \ldots, e_n$ — какие-либо линейно-независимые векторы пространства. Пусть при преобразовании они умножаются на некоторые числа $\lambda_1, \lambda_2, \ldots, \lambda_n$. Если векторы $e_1, e_2, \ldots, e_n$ принять за базис пространства, то рассматриваемое преобразование описывается диагональной матрицей

$$\begin{pmatrix} \lambda_1 & 0 & \ldots & 0 \\ 0 & \lambda_2 & \ldots & 0 \\ \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \ldots & \lambda_n \end{pmatrix}.$$

Преобразования этого класса имеют простой и наглядный геометрический смысл (конечно, для действительных пространств и при $n=2$ или $n=3$). Именно, если все числа λ_i положительны, то описываемое преобразование заключается в растяжении (или сжатии) пространства по направлению векторов $e_1, e_2, \ldots, e_n$ с коэффициентами $\lambda_1, \lambda_2, \ldots, \lambda_n$. Если некоторые из λ_i отрицательны, то деформация пространства сопровождается изменением направлений некоторых из векторов $e_1, e_2, \ldots, e_n$ на противоположные. Наконец, если, например, $\lambda_1=0$, то происходит проектирование пространства параллельно e_1 на подпространство, натянутое на $e_2, \ldots, e_n$ с последующей деформацией по этим направлениям.

Рассматриваемый класс преобразований важен тем, что он, несмотря на свою простоту, является весьма общим. Именно, устанавливается, что каждое линейное преобразование, удовлетворяющее некоторым, не очень жестким ограничениям, принадлежит к рассматриваемому классу, т. е. для него можно найти такой базис, в котором оно описывается диагональной матрицей.

Ограничения, накладываемые на преобразование, становятся особенно свободными, если рассматривать линейные преобразования комплексного пространства. В дальнейшем это и будет предполагаться.

Введем следующее определение.

Ненулевой вектор X, переходящий при линейном преобразовании A пространства в коллинеарный вектор λX, называется *собственным вектором преобразования*. Иными словами, ненулевой вектор X есть собственный вектор преобразования A в том и только в том случае, если $AX=\lambda X$. Число же λ называется *собственным значением* преобразования A.

Очевидно, что если преобразование имеет в каком-либо базисе диагональную матрицу, то этот базис состоит из собственных векторов, а сами диагональные элементы являются собственными значениями. Обратно, если в пространстве существует базис, состоящий из собственных векторов преобразования A, то в этом базисе матрица преобразования A будет диагональной и составленной из собственных значений, соответствующих векторам базиса.

Перейдем к изучению свойств собственных векторов и собственных значений. С этой целью перейдем к координатной записи в определении собственного вектора. Пусть. A есть матрица, соответствующая преобразованию $\boldsymbol{A}$ относительно некоторого базиса, X — столбец из координат вектора $\boldsymbol{X}$ в том же базисе. Равенство $\boldsymbol{AX}=\lambda\boldsymbol{X}$ при переходе к координатам запишется как $AX=\lambda X$ или

$$(A-\lambda E)X=0.$$

В развернутой форме это равенство превращается в систему

$$\begin{aligned}
&(a_{11}-\lambda)x_1+a_{12}x_2+\ldots+a_{1n}x_n=0,\\
&a_{21}x_1+(a_{22}-\lambda)x_2+\ldots+a_{2n}x_n=0,\\
&\ldots\ldots\ldots\ldots\ldots\ldots\ldots\\
&a_{n1}x_1+a_{n2}x_2+\ldots+(a_{nn}-\lambda)x_n=0.
\end{aligned}$$

Эту систему равенств можно рассматривать как систему линейных однородных уравнений относительно $x_1, x_2, \ldots, x_n$. Нас интересует случай, когда эта система имеет нетривиальное решение, ибо координаты собственного вектора не должны равняться нулю одновременно. Мы знаем, что необходимым и достаточным условием существования нетривиального решения системы линейных однородных уравнений является требование, чтобы ранг матрицы из коэффициентов был меньше числа неизвестных, а это равносильно тому, что определитель системы равен нулю

$$\begin{vmatrix}
a_{11}-\lambda & a_{12} & \ldots & a_{1n}\\
a_{21} & a_{22}-\lambda & \ldots & a_{2n}\\
\ldots & \ldots & \ldots & \ldots\\
a_{n1} & a_{n2} & \ldots & a_{nn}-\lambda
\end{vmatrix}=0.$$

Таким образом, все собственные числа преобразования $\boldsymbol{A}$ являются корнями многочлена $|A-\lambda E|$ и, обратно, каждый корень этого многочлена является собственным значением преобразования, так как каждому корню соответствует по крайней мере один собственный вектор. Многочлен $|A-\lambda E|$ называется *характеристическим многочленом* матрицы A. Уравнение же $|A-\lambda E|=0$ называется *характеристическим* или *вековым уравнением* матрицы A, а его корни — *характеристическими числами* матрицы[1].

По основной теореме высшей алгебры (глава IV, том 1) каждый многочлен имеет по крайней мере один корень, следовательно, каждое линейное преобразование имеет по крайней мере одно собственное число и, следовательно, по крайней мере один собственный вектор. Но, конечно, при этом возможно, что, даже в случае, если преобразование описывается действительной матрицей, может оказаться, что все или

[1] Название «вековое уравнение» возникло в небесной механике в связи с задачей о так называемых вековых возмущениях в движении планет.

часть его собственных чисел комплексны. И на самом деле, в действительном пространстве теорема о существовании (действительных) собственных чисел и собственных векторов для любого линейного преобразования оказывается неверной. Например, преобразование плоскости, заключающееся в повороте вокруг начала координат на некоторый угол, отличный от 180°, изменяет направления всех векторов плоскости, так что собственных векторов для этого преобразования не существует.

Корни характеристического многочлена матрицы A являются собственными числами преобразования $\boldsymbol{A}$, и, следовательно, матрицы, отвечающие одному и тому же преобразованию в различных базисах, обладают одинаковыми совокупностями корней характеристического многочлена. Это делает правдоподобным предположение, что и сам характеристический многочлен линейного преобразования зависит только от преобразования, но не от выбора базиса. Это проверяется следующей изящной выкладкой, основанной на свойствах действий над матрицами и определителями.

Мы знаем, что если матрица A соответствует преобразованию $\boldsymbol{A}$ в некотором базисе, то в каком-либо другом базисе преобразование $\boldsymbol{A}$ имеет подобную матрицу $C^{-1}AC$, где C — некоторая неособенная матрица. Но

$$|C^{-1}AC-\lambda E|=|C^{-1}AC-C^{-1}\lambda EC|=|C^{-1}(A-\lambda E)C|=$$
$$=|C^{-1}||C||A-\lambda E|=|C^{-1}C||A-\lambda E|=|A-\lambda E|.$$

Таким образом, матрицы, соответствующие одному и тому же преобразованию $\boldsymbol{A}$ в различных базисах, действительно имеют один и тот же характеристический многочлен, который может быть поэтому назван характеристическим многочленом преобразования.

Сделаем теперь предположение, что все собственные числа преобразования $\boldsymbol{A}$ различны. Докажем, что собственные векторы, взятые по одному для каждого собственного числа, линейно-независимы. Действительно, если допустить, что некоторые из них, именно $e_1, \ldots, e_k$, линейно-независимы, а остальные, в том числе и e_{k+1}, являются их линейными комбинациями, то

$$e_{k+1}=c_1e_1+c_2e_2+\ldots+c_ke_k. \tag{13}$$

Применив к обеим частям равенства линейное преобразование, получим

$$\boldsymbol{A}e_{k+1}=c_1\boldsymbol{A}e_1+c_2\boldsymbol{A}e_2+\ldots+c_k\boldsymbol{A}e_k,$$

откуда, в силу определения собственного вектора, следует, что

$$\lambda_{k+1}e_{k+1}=c_1\lambda_1e_1+c_2\lambda_2e_2+\ldots+c_k\lambda_ke_k.$$

Умножив равенство (13) на λ_{k+1} и вычтя из него вновь полученное равенство, будем иметь

$$c_1(\lambda_{k+1}-\lambda_1)e_1+c_2(\lambda_{k+1}-\lambda_2)e_2+\ldots+c_k(\lambda_{k+1}-\lambda_k)e_k=0.$$

Отсюда следует, в силу линейной независимости $e_1, e_2, \ldots, e_k$, что

$$c_1(\lambda_{k+1}-\lambda_1)=c_2(\lambda_{k+1}-\lambda_2)=\ldots=c_k(\lambda_{k+1}-\lambda_k)=0.$$

Но мы предположили, что все собственные числа различны и векторы взяты по одному для каждого собственного числа. Следовательно, $\lambda_{k+1}-\lambda_1\neq 0$, $\lambda_{k+1}-\lambda_2\neq 0, \ldots, \lambda_{k+1}-\lambda_k\neq 0$, и равенство (13) неосуществимо, так как коэффициенты $c_1, c_2, \ldots, c_k$ не могут одновременно равняться нулю.

Теперь ясно, что если все собственные числа линейного преобразования различны, то существует базис, в котором матрица преобразования имеет диагональный вид. Действительно, за такой базис можно взять систему собственных векторов, взятых по одному для каждого собственного числа. Как мы доказали, они линейно-независимы, и их число равно числу измерений пространства, т. е. они действительно образуют базис.

Доказанная теорема в терминах теории матриц формулируется так. Если все собственные числа матрицы различны, то матрица подобна диагональной, диагональными элементами которой являются эти собственные числа.

Вопрос о преобразовании матрицы линейного преобразования к простейшему виду, в случае если среди корней характеристического многочлена имеются равные, значительно сложнее. Ограничимся кратким описанием окончательного результата.

«Каноническим ящиком» порядка m называется матрица вида

$$I_{m,\lambda_i}=\begin{pmatrix}\lambda_i & 1 & & & \\ & \lambda_i & 1 & & \\ & & \ddots & \ddots & \\ & & & \ddots & 1 \\ & & & & \lambda_i\end{pmatrix}.$$

Все необозначенные элементы равны нулю.

Канонической матрицей Жордана называется матрица, вдоль главной диагонали которой расположены «канонические ящики», а все остальные элементы равны нулю:

$$\begin{pmatrix}I_{m_1,\lambda_1} & & & \\ & I_{m_2,\lambda_2} & & \\ & & \ddots & \\ & & & I_{m_k,\lambda_k}\end{pmatrix}.$$

Числа λ_i в различных «ящиках» не обязаны быть попарно различными. Любая матрица может быть приведена к подобной ей канонической матрице Жордана. Доказательство этой теоремы довольно сложно. Необходимо заметить, что эта теорема играет большую роль во многих приложениях алгебры к другим вопросам математики, в частности в теории систем линейных дифференциальных уравнений.

Матрица приводится к диагональной форме в том и только в том случае, если порядки m_i всех ящиков равны единице.

§ 5. КВАДРАТИЧНЫЕ ФОРМЫ

Определение и простейшие свойства. *Квадратичной формой* называется однородный многочлен 2-й степени от нескольких переменных.

Квадратичная форма от n переменных $x_1, x_2, \ldots, x_n$ состоит из слагаемых двух типов: квадратов переменных и их попарных произведений с некоторыми коэффициентами. Квадратичную форму принято записывать в виде следующей квадратной схемы:

$$\begin{aligned} f(x_1, x_2, \ldots, x_n) = {} & a_{11}x_1^2 + a_{12}x_1x_2 + \ldots + a_{1n}x_1x_n + \\ & + a_{21}x_2x_1 + a_{22}x_2^2 + \ldots + a_{2n}x_2x_n + \\ & \ldots\ldots\ldots\ldots\ldots\ldots \\ & + a_{n1}x_nx_1 + a_{n2}x_nx_2 + \ldots + a_{nn}x_n^2. \end{aligned}$$

Пары подобных членов $a_{12}x_1x_2$ и $a_{21}x_2x_1$ и т. д. записываются с одинаковыми коэффициентами, так что каждый из них составляет половину коэффициента при соответствующем произведении переменных. Таким образом, каждая квадратичная форма естественным образом связывается с матрицей ее коэффициентов, которая является симметричной.

Квадратичную форму удобно представлять и в следующей матричной записи. Обозначим через X столбец из переменных $x_1, x_2, \ldots, x_n$, через $\overline{X}$ — строку $(x_1, x_2, \ldots, x_n)$, т. е. матрицу, транспонированную с X. Тогда

$$\begin{aligned} & f(x_1, x_2, \ldots, x_n) = x_1(a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n) + \\ & + x_2(a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n) + \ldots + x_n(a_{n1}x_1 + a_{n2}x_2 + \ldots + a_{nn}x_n) = \\ & = (x_1, x_2, \ldots, x_n)\begin{pmatrix} a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n \\ a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n \\ \ldots\ldots\ldots\ldots\ldots \\ a_{n1}x_1 + a_{n2}x_2 + \ldots + a_{nn}x_n \end{pmatrix} = \\ & = (x_1, x_2, \ldots, x_n)\begin{pmatrix} a_{11}a_{12}\ldots a_{1n} \\ a_{21}a_{22}\ldots a_{2n} \\ \ldots\ldots\ldots \\ a_{n1}a_{n2}\ldots a_{nn} \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \overline{X}AX. \end{aligned}$$

Квадратичные формы встречаются во многих разделах математики и ее приложений.

В теории чисел и кристаллографии рассматриваются квадратичные формы в предположении, что переменные $x_1, x_2, \ldots, x_n$ принимают только целочисленные значения. В аналитической геометрии квадратичная форма входит в состав уравнения кривой (или поверхности) 2-го порядка. В механике и физике квадратичная форма появляется для выражения кинетической энергии системы через компоненты обобщенных скоростей и т. д. Но, кроме того, изучение квадратичных форм необходимо и в анализе при изучении функций от многих переменных, в вопросах, для решения которых важно выяснить, как данная функция в окрестности данной точки отклоняется от приближающей ее линейной функции. Примером задачи этого типа является исследование функции на максимум и минимум.

Рассмотрим, например, задачу об исследовании на максимум и минимум для функции от двух переменных $w = f(x, y)$, имеющей непрерывные частные производные до 3-го порядка. Необходимым условием для того, чтобы точка (x_0, y_0) давала максимум или минимум функции w, является равенство нулю частных производных 1-го порядка в точке (x_0, y_0). Допустим, что это условие выполнено. Придадим переменным x и y малые приращения h и k и рассмотрим соответствующее приращение функции $\Delta w = f(x_0 + h, y_0 + k) - f(x_0, y_0)$. Согласно формуле Тейлора это приращение с точностью до малых высших порядков равно квадратичной форме $\frac{1}{2}(rh^2 + 2shk + tk^2)$, где r, s и t — значения вторых производных $\frac{\partial^2 w}{\partial x^2}$, $\frac{\partial^2 w}{\partial x \partial y}$, $\frac{\partial^2 w}{\partial y^2}$, вычисленные в точке (x_0, y_0). Если эта квадратичная форма положительна при всех значениях h и k (кроме $h = k = 0$), то функция w имеет минимум в точке (x_0, y_0); если отрицательна, то — максимум. Наконец, если форма принимает и положительные и отрицательные значения, то не будет ни максимума, ни минимума. Аналогичным образом исследуются и функции от большего числа переменных.

Изучение квадратичных форм в основном заключается в исследовании проблемы эквивалентности форм относительно той или другой совокупности линейных преобразований переменных. Две квадратичные формы называются *эквивалентными*, если одна из них может быть переведена в другую посредством одного из преобразований данной совокупности. С проблемой эквивалентности тесно связана проблема *приведения* формы, т. е. преобразования ее к некоторому возможно простейшему виду.

В различных вопросах, связанных с квадратичными формами, рассматриваются и различные совокупности допустимых преобразований переменных.

В вопросах анализа применяются любые неособенные преобразования переменных; для целей аналитической геометрии наибольший интерес представляют ортогональные преобразования, т. е. те, которым соответствует переход от одной системы переменных декартовых координат к другой. Наконец, в теории чисел и в кристаллографии рассматриваются линейные преобразования с целыми коэффициентами и с определителем, равным единице.

Мы рассмотрим из этих задач две: вопрос о приведении квадратичной формы к простейшему виду посредством любых неособенных преобразований и тот же вопрос для преобразований ортогональных. Прежде всего выясним, как преобразуется матрица квадратичной формы при линейном преобразовании переменных.

Пусть $f(x_1, x_2, \ldots, x_n) = \overline{X}AX$, где A — симметричная матрица из коэффициентов формы, X — столбец из переменных.

Сделаем линейное преобразование переменных, записав его сокращенно $X = CX'$. Здесь C обозначает матрицу коэффициентов этого преобразования, X' — столбец из новых переменных. Тогда $\overline{X} = \overline{X}'\overline{C}$, и, следовательно, $\overline{X}AX = \overline{X}'(\overline{C}AC)X'$, так что матрицей преобразованной квадратичной формы является $\overline{C}AC$.

Матрица $\overline{C}AC$ автоматически оказывается симметричной, что легко проверяется. Таким образом, задача о приведении квадратичной формы к простейшему виду равносильна задаче о приведении к простейшему виду симметричной матрицы посредством умножения ее слева и справа на взаимно транспонированные матрицы.

Преобразование квадратичной формы к каноническому виду посредством последовательного выделения квадратов. Установим, что любая (действительная) квадратичная форма может быть приведена к сумме квадратов новых переменных с некоторыми коэффициентами посредством действительного неособенного линейного преобразования.

Для доказательства установим прежде всего, что если форма не равна нулю тождественно, то за счет неособенного преобразования переменных можно сделать коэффициент при квадрате первой переменной отличным от нуля.

В самом деле, пусть

$$\begin{aligned} f(x_1, x_2, \ldots, x_n) = {} & a_{11}x_1^2 + a_{12}x_1x_2 + \ldots + a_{1n}x_1x_n + \\ & + a_{21}x_2x_1 + a_{22}x_2^2 + \ldots + a_{2n}x_2x_n + \\ & \ldots\ldots\ldots\ldots\ldots\ldots \\ & + a_{n1}x_nx_1 + a_{n2}x_nx_2 + \ldots + a_{nn}x_n^2. \end{aligned}$$

Если $a_{11} \neq 0$, то никаких преобразований не требуется. Если $a_{11} = 0$, но какой-либо из диагональных коэффициентов $a_{kk} \neq 0$, то положим $x_1 = x_k'$, $x_k = x_1'$, приравнивая остальные исходные переменные соответ-

ствующим новым. Это неособенное преобразование приведет к цели. Наконец, если все диагональные коэффициенты равны нулю, то хотя бы один недиагональный коэффициент, например a_{12}, отличен от нуля.

Сделав неособенное преобразование

$$x_1 = x_1',$$
$$x_2 = x_1' + x_2'$$

и приравняв остальные исходные переменные к новым, мы достигаем цели.

Таким образом, без нарушения общности можно принять $a_{11} \neq 0$.

Выделим теперь квадрат линейной функции так, чтобы все слагаемые, содержащие x_1, вошли в этот квадрат.

Это легко сделать. Действительно,

$$\begin{aligned} f(x_1, x_2, \ldots, x_n) = {} & a_{11}x_1^2 + a_{12}x_1x_2 + \ldots + a_{1n}x_1x_n + \\ & + a_{21}x_2x_1 + a_{22}x_2^2 + \ldots + a_{2n}x_2x_n + \\ & \ldots\ldots\ldots\ldots\ldots\ldots \\ & + a_{n1}x_nx_1 + a_{n2}x_nx_2 + \ldots + a_{nn}x_n^2 = \\ = {} & a_{11}\left(x_1 + \frac{a_{12}}{a_{11}}x_2 + \ldots + \frac{a_{1n}}{a_{11}}x_n\right)^2 - a_{11}\left(\frac{a_{12}}{a_{11}}x_2 + \ldots + \frac{a_{1n}}{a_{11}}x_n\right)^2 + \\ & + a_{22}x_2^2 + \ldots + a_{2n}x_2x_n + \\ & \ldots\ldots\ldots\ldots \\ & + a_{n2}x_nx_2 + \ldots + a_{nn}x_n^2. \end{aligned}$$

Раскрыв скобки во втором слагаемом и сделав приведение подобных членов, получим

$$f(x_1, x_2, \ldots, x_n) = a_{11}\left(x_1 + \frac{a_{12}}{a_{11}}x_2 + \ldots + \frac{a_{1n}}{a_{11}}x_n\right)^2 + f_1(x_2, \ldots, x_n),$$

где f_1 есть форма уже от $n-1$ переменных.

Преобразование

$$\begin{aligned} x_1 + \frac{a_{12}}{a_{11}}x_2 + \ldots + \frac{a_{1n}}{a_{11}}x_n &= x_1', \\ x_2 &= x_2', \\ \ldots\ldots & \\ x_n &= x_n', \end{aligned}$$

очевидно, неособенное. Сделав это преобразование, приведем форму к виду

$$a_{11}x_1'^2 + f_1(x_2', \ldots, x_n').$$

Продолжая процесс аналогичным образом, мы приведем форму к требуемому «каноническому» виду

$$\alpha_1 z_1^2 + \alpha_2 z_2^2 + \ldots + \alpha_n z_n^2.$$

Здесь $z_1, z_2, \ldots, z_n$ — последние из введенных новых переменных.

Закон инерции квадратичных форм. При приведении квадратичной формы к каноническому виду имеется весьма значительный произвол в выборе осуществляющего это приведение преобразования переменных. Этот произвол виден хотя бы из того, что имеется возможность раньше, чем применить изложенный выше способ последовательного выделения квадратов, сделать любое неособенное преобразование переменных.

Однако, несмотря на этот произвол, в результате приведения данной формы получаются почти одинаковые канонические квадратичные формы независимо от выбора приводящего преобразования. Именно, число квадратов новых переменных, входящих с положительными, отрицательными и нулевыми коэффициентами, получается одним и тем же при любом способе приведения. Эта теорема носит название *закона инерции* квадратичных форм. На его доказательстве мы не будем останавливаться.

Закон инерции квадратичной формы решает задачу об эквивалентности действительной квадратичной формы относительно всех неособенных преобразований. Именно, две формы эквивалентны в том и только в том случае, если при приведении их к каноническому виду получаются канонические формы с одинаковым числом квадратов с положительными, отрицательными и нулевыми коэффициентами.

Особый интерес для приложений имеют квадратичные формы, которые после приведения к каноническому виду превращаются в сумму квадратов новых переменных со всеми *положительными* коэффициентами. Такие формы носят название *положительно-определенных*.

Положительно-определенные квадратичные формы характеризуются тем свойством, что все значения их при действительных значениях переменных, не равных одновременно нулю, положительны.

Ортогональное преобразование квадратичных форм к каноническому виду. Среди всевозможных способов приведения квадратичной формы к каноническому виду особый интерес представляют ортогональные преобразования, т. е. осуществляющиеся посредством линейного преобразования переменных с ортогональной матрицей. Именно такие преобразования представляют интерес, например, в аналитической геометрии — в задаче о приведении общего уравнения кривой или поверхности 2-го порядка к каноническому виду.

Для того, чтобы убедиться в возможности такого преобразования, целесообразно рассматривать квадратичную форму как функцию от вектора в эвклидовом пространстве, рассматривая переменные $x_1, x_2, \ldots, x_n$

как координаты переменного вектора относительно некоторого ортонормального базиса. Тогда ортогональное преобразование переменных интерпретируется как переход от одного ортонормального базиса к другому.

Свяжем с квадратичной формой

$$f(x_1, x_2, \ldots, x_n) = a_{11}x_1^2 + \ldots + a_{1n}x_1x_n + \\ \cdots\cdots\cdots \\ + a_{n1}x_nx_1 + \ldots + a_{nn}x_n^2$$

линейное преобразование A, имеющее по отношению к выбранному базису матрицу $A = \begin{pmatrix} a_{11} \ldots a_{1n} \\ \ldots\ldots \\ a_{n1} \ldots a_{nn} \end{pmatrix}$. Тогда сама квадратичная форма может рассматриваться как скалярное произведение $AX \cdot X$ (где X — вектор с координатами $x_1, x_2, \ldots, x_n$), а ее коэффициенты a_{ij} — как скалярные произведения $Ae_i \cdot e_j$, где $e_1, e_2, \ldots, e_n$ — выбранный ортонормальный базис.

Легко видеть, что вследствие симметрии матрицы A для любых векторов X и Y имеет место равенство

$$AX \cdot Y = X \cdot AY.$$

Докажем прежде всего, что преобразование A имеет по крайней мере одно действительное собственное число и соответствующий ему собственный вектор.

Для этого рассмотрим значения формы $AX \cdot X$ в предположении, что вектор X пробегает единичную сферу, т. е. совокупность всех единичных векторов. При этих условиях форма $AX \cdot X$ будет иметь максимум. Покажем, что этот максимум λ_1 есть собственное число преобразования A, а вектор X_0, для которого этот максимум достигается, есть соответствующий собственный вектор, т. е. $AX_0 = \lambda_1 X_0$.

Доказательство этого утверждения проведем косвенными средствами, установив, что вектор AX_0 ортогонален ко всем векторам, ортогональным к X_0.

Заметим, что для любого вектора Z справедливо неравенство $AZ \cdot Z \leqslant \lambda_1 |Z|^2$. Это очевидно из того, что $X = \frac{Z}{|Z|}$ есть единичный вектор, а λ_1 — максимум значения формы $AX \cdot X$ на единичной сфере. Рассмотрим $Z = X_0 + \varepsilon Y$, где ε — некоторое действительное число, Y — произвольный вектор, ортогональный к вектору X_0. Тогда

$$AZ \cdot Z = (AX_0 + \varepsilon AY) \cdot (X_0 + \varepsilon Y) = AX_0 \cdot X_0 + 2\varepsilon AX_0 \cdot Y + \varepsilon^2 AY \cdot Y = \\ = \lambda_1 + 2\varepsilon AX_0 \cdot Y + \varepsilon^2 AY \cdot Y.$$

Кроме того,

$$|Z|^2=(X_0+\varepsilon Y)\cdot(X_0+\varepsilon Y)=|X_0|^2+\varepsilon^2|Y|^2=1+\varepsilon^2|Y|^2,$$

ибо

$$X_0\cdot Y=0,\quad |X_0|^2=1.$$

Следовательно,

$$\lambda_1+2\varepsilon AX_0\cdot Y+\varepsilon^2 AY\cdot Y\leqslant\lambda_1+\varepsilon^2\lambda_1|Y|^2,$$

откуда, поделив на ε^2, получим

$$\frac{2}{\varepsilon}AX_0\cdot Y\leqslant\lambda_1|Y|^2-AY\cdot Y. \tag{14}$$

Последнее неравенство должно выполняться при любом действительном ε, сколь угодно малом по абсолютной величине.

Но оно может выполняться только при условии $AX_0\cdot Y=0$, так как если $AX_0\cdot Y>0$, то неравенство (14) невозможно при достаточно малом положительном ε, если же $AX_0\cdot Y<0$, то оно невозможно при достаточно малом по абсолютной величине отрицательном ε. Итак, $AX_0\cdot Y=0$, т. е. AX_0 действительно ортогонален ко всякому вектору, ортогональному к X_0. Следовательно, AX_0 и X_0 коллинеарны, т. е. $AX_0=\lambda' X_0$, где λ' — некоторое действительное число. То, что $\lambda'=\lambda_1$, легко проверить, именно

$$\lambda_1=AX_0\cdot X_0=\lambda' X_0\cdot X_0=\lambda'.$$

Теперь легко доказать, что каждая квадратичная форма действительно может быть приведена к каноническому виду посредством ортогонального преобразования.

Пусть $e_1, e_2, \ldots, e_n$ — исходный ортонормальный базис пространства, $f_1, f_2, \ldots, f_n$ — новый ортонормальный базис, в котором первый вектор f_1 равен собственному вектору X_0 преобразования A. Пусть $x_1, x_2, \ldots, x_n$ — координаты вектора X в исходном базисе, а $x'_1, x'_2, \ldots, x'_n$ — его же координаты в новом базисе. Тогда

$$\begin{pmatrix} x_1\\ x_2\\ \cdot\\ \cdot\\ \cdot\\ x_n\end{pmatrix}=P\begin{pmatrix} x'_1\\ x'_2\\ \cdot\\ \cdot\\ \cdot\\ x'_n\end{pmatrix}$$

где P — ортогональная матрица.

Сделаем в квадратичной форме $AX\cdot X$ переход к новым переменным. В новых переменных квадратичная форма будет иметь коэффициенты $a'_{ij}=Af_i\cdot f_j$. Следовательно,

$$a'_{11}=Af_1\cdot f_1=\lambda_1 f_1\cdot f_1=\lambda_1,$$

$$a'_{1j}=a'_{j1}=Af_1\cdot f_j=\lambda_1 f_1\cdot f_j=0 \text{ при } j\neq 1,$$

т. е. форма имеет вид

$$\lambda_1 x_1'^2 + \varphi(x_2', \ldots, x_n').$$

Итак, за счет ортогонального преобразования нам удалось выделить один квадрат новой переменной.

Проведя те же рассуждения с новой формой $\varphi(x_2', \ldots, x_n')$ и т. д., мы придем в конце концов к тому, что форма посредством цепочки ортогональных преобразований окажется приведенной к каноническому виду. Но очевидно, что цепочка ортогональных преобразований равносильна одному ортогональному же преобразованию. Тем самым теорема доказана.

§ 6. ФУНКЦИИ ОТ МАТРИЦ И НЕКОТОРЫЕ ИХ ПРИЛОЖЕНИЯ

Функции от матриц. Приложения линейной алгебры к другим отделам математики весьма многочисленны и разнообразны. Не будет преувеличением сказать, что в большей части современной математики и теоретической физики в той или иной форме используются идеи и результаты линейной алгебры, главным образом в форме исчисления матриц.

Рассмотрим коротко один из путей приложения исчисления матриц к теории обыкновенных дифференциальных уравнений. Здесь важную роль играют функции от матриц.

Прежде всего определим степень квадратной матрицы A. Положим $A^0 = E$, $A^1 = A$, $A^2 = AA$, $A^3 = A^2A$, $A^4 = A^3A$ и т. д. При помощи сочетательного закона легко доказать, что $A^m A^n = A^{m+n}$ для любых натуральных m и n. Для матриц были определены действия сложения и умножения на число. Это дает возможность естественным образом определить значение многочлена (от одной переменной) от матрицы. Именно, если $\varphi(x) = a_0 x^n + a_1 x^{n-1} + \ldots + a_n$, то положим (по определению) $\varphi(A) = a_0 A^n + a_1 A^{n+1} + \ldots + a_n E$. Таким образом определяется понятие наиболее простой функции от матричного аргумента — многочлена.

Посредством предельного перехода легко обобщить понятие функции от матричного аргумента на значительно более широкий класс функций, чем многочлен от одной переменной. Не касаясь этого вопроса во всей его общности, ограничимся рассмотрением аналитических функций.

Прежде всего введем понятие предела последовательности матриц. Последовательность матриц

$$A_1 = \begin{pmatrix} a_{11}^{(1)} \ldots a_{1n}^{(1)} \\ \cdots\cdots \\ a_{n1}^{(1)} \ldots a_{nn}^{(1)} \end{pmatrix}, \quad A_2 = \begin{pmatrix} a_{11}^{(2)} \ldots a_{1n}^{(2)} \\ \cdots\cdots \\ a_{n1}^{(2)} \ldots a_{nn}^{(2)} \end{pmatrix}, \ldots$$

называется *сходящейся* к матрице $A = \begin{pmatrix} a_{11} \ldots a_{1n} \\ \cdots\cdots \\ a_{n1} \ldots a_{nn} \end{pmatrix}$ (или имеющей пре-

делом матрицу A), если $\lim\limits_{k\to\infty} a_{ij}^{(k)} = a_{ij}$ для всех i, j. Далее, суммой ряда $A_1 + A_2 + \ldots + A_k + \ldots$ называется предел сумм его отрезков $\lim\limits_{k\to\infty}(A_1 + A_2 + \ldots + A_k)$, если этот предел существует.

Пусть $f(z)$ есть аналитическая функция, регулярная в окрестности $z = 0$. Тогда, как известно, $f(z)$ разлагается в степенной ряд

$$f(z) = a_0 + a_1 z + a_2 z^2 + \ldots + a_k z^k + \ldots$$

Для любой квадратной матрицы A естественно положить

$$f(A) = a_0 E + a_1 A + a_2 A^2 + \ldots + a_k A^k + \ldots$$

Оказывается, что такой ряд сходится для всех матриц A, собственные числа которых лежат внутри круга сходимости степенного ряда $a_0 + a_1 z + \ldots + a_k z^k + \ldots$

В приложениях представляют интерес элементарные функции от матриц.

Так, например, геометрическая прогрессия $E + A + A^2 + \ldots + A^k + \ldots$ является сходящимся рядом для матриц, модули собственных чисел которых меньше 1, и суммой этого ряда является матрица $(E - A)^{-1}$, что находится в полном соответствии с формулой

$$1 + x + \ldots + x^k + \ldots = \frac{1}{1-x}.$$

Представление $(E - A)^{-1}$ в виде бесконечного ряда дает эффективное средство для приближенного решения систем линейных уравнений, матрицы коэффициентов которых близки к единичной.

Действительно, записав такую систему в форме

$$(E - A)X = B,$$

получим

$$X = (E - A)^{-1} B = B + AB + A^2 B + \ldots, \tag{15}$$

что дает удобную формулу для решения системы, если только ряд (15) сходится достаточно быстро.

Полезно рассматривать биномиальный ряд

$$(E + A)^m = E + \frac{m}{1} A + \frac{m(m-1)}{2!} A^2 + \ldots,$$

который можно применять (если собственные числа A по модулю меньше 1) не только для натуральных показателей m, но и для дробных и отрицательных.

Особенно важной для приложений является показательная функция от матрицы

$$e^A = E + A + \frac{A^2}{2!} + \frac{A^3}{3!} + \ldots$$

Ряд, определяющий показательную функцию, сходится при любой матрице A. Показательная функция от матриц обладает свойствами, напоминающими свойства обычной показательной функции. Так, если A и B коммутируют при умножении, т. е. $AB=BA$, то $e^{A+B}=e^A\cdot e^B$. Однако при некоммутирующих A и B формула перестает быть верной.

Приложение к теории систем обыкновенных линейных дифференциальных уравнений. В теории систем обыкновенных линейных дифференциальных уравнений целесообразно рассматривать матрицы, элементы которых являются функциями от некоторой независимой переменной:

$$U(t)=\begin{pmatrix} a_{11}(t)\ldots a_{1n}(t) \\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot \\ a_{m1}(t)\ldots a_{mn}(t) \end{pmatrix}.$$

Для таких матриц естественным образом определяется понятие производной по аргументу t. Именно:

$$\frac{dU(t)}{dt}=\begin{pmatrix} a'_{11}(t)\ldots a'_{1n}(t) \\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot \\ a'_{m1}(t)\ldots a'_{mn}(t) \end{pmatrix}.$$

Нетрудно проверить, что для матриц справедливы некоторые элементарные формулы дифференцирования. Так

$$\frac{d(U+V)}{dt}=\frac{dU}{dt}+\frac{dV}{dt},$$

$$\frac{d(cU)}{dt}=c\frac{dU}{dt},$$

$$\frac{d(UV)}{dt}=\frac{dU}{dt}V+U\frac{dV}{dt}.$$

(Умножать нужно строго в том порядке, какой дан в формуле!)

Система обыкновенных линейных однородных дифференциальных уравнений

$$\frac{dy_1}{dt}=a_{11}(t)y_1+a_{12}(t)y_2+\ldots+a_{1n}(t)y_n,$$

$$\frac{dy_2}{dt}=a_{21}(t)y_1+a_{22}(t)y_2+\ldots+a_{2n}(t)y_n,$$

$$\cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot$$

$$\frac{dy_n}{dt}=a_{n1}(t)y_1+a_{n2}(t)y_2+\ldots+a_{nn}(t)y_n$$

в этих обозначениях может быть записана в виде

$$\frac{dY}{dt}=A(t)Y,$$

где

$$Y=\begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}, \quad A(t)=\begin{pmatrix} a_{11}(t)\ldots a_{1n}(t) \\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot \\ a_{n1}(t)\ldots a_{nn}(t) \end{pmatrix},$$

т. е. в форме, аналогичной одному линейному однородному дифференциальному уравнению.

Если коэффициенты системы постоянные, т. е. матрица A постоянная, то и решение системы внешне выглядит так же, как решение уравнения $y'=ay$. Именно, в этом случае $Y=e^{At}C$, где C — столбец из произвольных постоянных.

Решение в этой форме очень удобно для исследования. Дело в том, что для любой аналитической функции $f(z)$ имеет место равенство

$$f(B^{-1}LB)=B^{-1}f(L)B.$$

Так как любую матрицу можно привести к канонической форме Жордана (см. § 4), то вычисление функции от любой матрицы сводится к вычислению функции от канонической матрицы, что легко осуществить. Поэтому если $A=B^{-1}LB$, где L — каноническая матрица, то

$$Y=e^{At}C=B^{-1}e^{Lt}BC=B^{-1}e^{Lt}C',$$

где $C'=BC$ — столбец из произвольных постоянных.

Из этой формулы нетрудно получить явное выражение для всех составляющих искомого столбца Y.

Советский ученый И. А. Лаппо-Данилевский с успехом развил аппарат теории функций от матриц и впервые применил его к исследованию систем также и с переменными коэффициентами. Его результаты принадлежат к числу наиболее блестящих достижений математики за последние пятьдесят лет.

ЛИТЕРАТУРА

Гантмахер Ф. Р. Теория матриц. Гостехиздат, 1953.
Монография, содержащая богатый материал по приложениям линейной алгебры.

Гельфанд И. М. Лекции по линейной алгебре. Гостехиздат, 1951.
Содержит геометрическое изложение линейной алгебры с выходом в функциональный анализ.

Курош А. Г. Курс высшей алгебры. Гостехиздат, 1952.

Мальцев А. И. Основы линейной алгебры. Изд. 2, Гостехиздат, 1956.

Окунев Л. Я. Высшая алгебра. Гостехиздат, 1949.

Смирнов В. И. Курс высшей математики, т. III. Гостехиздат, 1953.

Фаддеева В. Н. Вычислительные методы линейной алгебры. Гостехиздат, 1950.
Излагаются методы численного решения основных задач линейной алгебры.

Глава XVII

АБСТРАКТНЫЕ ПРОСТРАНСТВА

С тех пор как Н. И. Лобачевский впервые показал возможность неэвклидовой геометрии и выдвинул новое представление об отношении геометрии к материальной действительности, предмет геометрии, ее методы и применения чрезвычайно расширились. Теперь математики изучают разные «пространства»: наряду с эвклидовым пространством рассматривается пространство Лобачевского, проективное пространство, различные n-мерные и даже бесконечномерные пространства, римановы, топологические и другие пространства; число таких пространств неограничено, и каждое из них имеет свои свойства, свою «геометрию». В физике используют понятия о так называемых фазовых и конфигурационных «пространствах»; теория относительности применяет представление о кривизне пространства и другие выводы абстрактных геометрических теорий.

Как и откуда возникли эти математические абстракции? Какое реальное основание, какое реальное значение и применение они имеют? Каково их отношение к действительности? Как они определяются и как их рассматривают в математике? Какое значение в математике имеют общие идеи современной геометрии?

На эти вопросы должна ответить настоящая глава. В ней не будут излагаться сами теории абстрактных математических пространств; это требовало бы гораздо большего объема изложения и гораздо большего обращения к специальному математическому аппарату. Задача состоит в том, чтобы выяснить сущность новых идей геометрии, т. е. ответить на поставленные вопросы, а это можно сделать без сложных доказательств и формул.

История вопроса восходит истоками к «Началам» Эвклида, к аксиоме или, как говорят еще, к постулату о параллельных линиях.

§ 1. ИСТОРИЯ ПОСТУЛАТА ЭВКЛИДА

В своих «Началах» Эвклид формулировал основные предпосылки геометрии в виде так называемых постулатов и аксиом. Среди них содержался V постулат (в других списках «Начал» — XI аксиома), который теперь формулируют обычно следующим образом: «Через точку, не лежа-

щую на данной прямой, нельзя провести более одной прямой, параллельной данной». Напомним, что прямая называется параллельной данной прямой, если обе прямые лежат в одной плоскости и не пересекаются, причем, говоря так, имеют в виду бесконечные прямые, а не конечные их отрезки.

Легко доказать, что через точку A, не лежащую на данной прямой a, всегда можно провести хотя бы одну прямую, параллельную данной.

Действительно, опустим из точки A перпендикуляр b на прямую a и проведем через A прямую c перпендикулярно b (рис. 1). Полученная фигура будет вполне симметрична относительно линии b, так как углы, образуемые прямой b по обе ее стороны с прямыми a и c, равны. Поэтому, перегибая плоскость по линии b, мы приведем половины прямых a и c в совпадение. Отсюда видно, что если бы a и c пересекались с одной стороны от AB, то они должны были бы пересекаться также с другой стороны. Выходило бы, что прямые a и c имеют две общие точки, а это невозможно, так как по основному свойству прямой через две точки может проходить только одна прямая (так что прямые, имеющие две общие точки, необходимо должны совпадать).

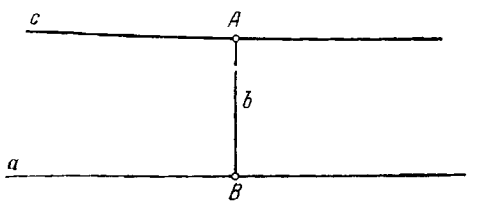


Рис. 1.

Итак, из основных свойств прямой и движения фигур (поскольку перегибание по линии AB есть вращение полуплоскости вокруг этой линии) следует, что хотя бы одна параллель к данной прямой всегда проходит через данную точку. Постулат же Эвклида дополняет этот вывод утверждением, что такая параллель только одна, никакой другой быть не может.

Среди других постулатов (аксиом) геометрии этот постулат занимает несколько особое место. У самого Эвклида он формулировался довольно сложно, но даже в приведенной выше обычной его форме он содержит известную трудность. Эта трудность заключается уже в самом понятии параллельных прямых: здесь речь идет о всей прямой. Но как убедиться, что данные прямые параллельны? Для этого нужно как бы пройти их в обе стороны «до бесконечности» и убедиться, что они нигде *на всем бесконечном протяжении* не пересекаются. Ясно, что такое представление имеет свою трудность. Все это, повидимому, и послужило причиной того, что постулат о параллельных занял уже у самого Эвклида несколько особое положение: в его «Началах» этот постулат применяется только начиная с 29-го предложения, в то время как в первых 28 предложениях Эвклид обходился без него. Ввиду сложности постулата желание обойтись без него могло возникнуть довольно естественно, и поэтому еще в древности появились попытки изменить определение параллельных линий, изме-

нить самую формулировку постулата или, что было бы лучше всего, вывести его как теорему из других аксиом и основных понятий геометрии.

Так, теория параллельных линий, основанная на V постулате, стала предметом комментирования и разработки в трудах многих геометров, начиная с древности. В цепи этих исследований главной задачей было вовсе избавиться от V постулата, выведя его как теорему из других основных положений геометрии.

Этой задачей занимались многие геометры: грек Прокл (V в. н. э.), комментировавший Эвклида, иранец Насирэддин Туси (XIII в.), англичанин Валлис (1616—1703), итальянец Саккери (1667—1733), немецкий философ и математик Ламберт (1728—1777), француз Лежандр (1752—1833) и многие другие; все они на протяжении более чем двух тысяч лет, прошедших со времени появления эвклидовых «Начал», изощрялись в тонкости и геометрическом остроумии, пытаясь доказать V постулат.

Однако результат этих попыток неизменно оставался отрицательным. Каждый раз выяснялось, что автор того или иного доказательства фактически опирался на какое-нибудь предположение, может быть, и очевидное, но вовсе не вытекающее с логической необходимостью из других предпосылок геометрии. Иными словами, дело сводилось каждый раз к замене V постулата другим утверждением, из которого этот постулат действительно вытекал, но которое само требовало доказательства [1].

Глубже других в задачу проникли Саккери и Ламберт. Саккери первый попытался доказать V постулат от противного, т. е. он принял за исходное противоположное утверждение и, развивая из него следствия, надеялся придти к противоречию. Дойдя в этих выводах до результатов, казавшихся совершенно невообразимыми, он подумал, что решил задачу. Но ошибся, потому что противоречие с наглядным представлением не означает еще логического противоречия. Задача ведь состояла в логическом доказательстве эвклидова постулата на основе других положений геометрии, а не в том, чтобы еще раз убедиться в его наглядной верности. Этот постулат и сам по себе наглядно достаточно убедителен. Но, повторяем, наглядная убедительность и логическая необходимость вещи различные.

Ламберт оказался более глубоким мыслителем, чем Саккери и его предшественники. Идя по тому же пути, он не нашел логического противоречия и не допустил ошибки других; он не заявил, будто доказал

[1] Таких утверждений, равносильных V постулату, было установлено много. Вот некоторые примеры: 1) прямая, параллельная данной, проходит от нее на постоянном расстоянии (Прокл); 2) существуют подобные (но не равные) треугольники, т. е. такие, углы которых равны, но стороны не равны (Валлис); 3) существует хотя бы один прямоугольник, т. е. такой четырехугольник, все углы которого прямые (Саккери); 4) прямая, перпендикулярная одной стороне острого угла, пересекает также другую его сторону (Лежандр); 5) сумма углов треугольника равна двум прямым (Лежандр); 6) существуют треугольники сколь угодно большой площади (Гаусс). Этот список теперь мы могли бы продолжать до бесконечности.

V постулат. Но и после него еще в начале XIX в. Лежандр снова «доказывает» V постулат, впадая в старую ошибку: постулат он опять-таки заменяет другими утверждениями, которые сами требуют доказательства.

Итак, к началу XIX в. проблема доказательства V постулата оставалась так же не решенной, как было во времена Эвклида. Усилия оставались тщетными, и задача, казалось, не продвинулась. Поистине то была глубокая загадка геометрии: задача, разрешимость которой казалась несомненной лучшим геометрам, никак не поддавалась решению в течение двух тысяч лет.

Теория параллельных стала в XIX в. одной из центральных задач геометрии. Ею занимались многие геометры: Гаусс, Лагранж, Даламбер, Лежандр, Вахтер, Швейкарт, Тауринус, Фаркаш Бóйаи и другие.

Однако доказательство постулата не удается. В чем же дело: в неумении ли решить задачу, или, может быть, задача неверно поставлена? Этот вопрос уже начинал возникать перед некоторыми из геометров, превосходившими других глубиной мысли. Гаусс, знаменитейший немецкий математик, бьется над задачей начиная с 1792 г. и постепенно перед ним вырисовывается правильная постановка вопроса. Наконец, он решается отказаться от V постулата и начиная с 1813 г. развивает последовательность теорем, выводимых из противоположного утверждения. Несколько позже тем же путем идут немецкие математики Швейкарт в бытность его профессором права в Харькове и затем Тауринус. Но никто из них не нашел окончательного ответа на вопрос. Гаусс тщательно скрывал свои исследования, Швейкарт ограничился частным письмом к Гауссу, и лишь Тауринус выступил в печати с элементами новой геометрии, основанной на отрицании V постулата. Однако он сам исключал возможность такой геометрии. Таким образом, никто из них не решил задачи, и вопрос о правильности всей ее постановки оставался без ответа. Ответ впервые был дан Н. И. Лобачевским, молодым профессором Казанского университета: 23 февраля 1826 г. он прочел в заседании физико-математического факультета доклад о теории параллельных, а в 1829 г. опубликовал его содержание в журнале Казанского университета.

§ 2. РЕШЕНИЕ ЛОБАЧЕВСКОГО

1. Сущность решения проблемы V постулата, данного Лобачевским, выражена им самим в сочинении «Новые начала геометрии» (1835) в следующих словах:

«Всем известно, что в геометрии теория параллельных до сих пор оставалась несовершенной. Напрасное старание со времен Эвклида, в продолжение двух тысяч лет, заставило меня подозревать, что в самих понятиях еще не заключается той истины, которую хотели доказывать и которую проверить, подобно другим физическим законам, могут лишь опыты, каковы, например, астрономические наблюдения. В справедли-

вости моей догадки будучи, наконец, убежден и почитая затруднительный вопрос решенным вполне, писал об этом я рассуждение в 1826 году».

Разберем, что имел в виду Лобачевский в этом высказывании, в котором, как в фокусе, сконцентрирована новая его идея, не только давшая решение вопроса о V постулате, но повернувшая по-новому все понимание геометрии, да и не одной геометрии.

Н. И. Лобачевский еще в 1815 г. начал работать над теорией параллельных, пытаясь вначале, подобно другим геометрам, доказать V постулат. В 1823 г. он уже ясно осознал, что все доказательства, «какие были даны, могут называться только пояснениями, но не заслуживают быть почтены в полном смысле математическими доказательствами» [1]. Тут он увидел, что «в самих понятиях не заключается той истины, которую хотели доказывать», т. е., иными словами, из основных посылок и понятий геометрии нельзя вывести V постулат. Как он убедился, что такой вывод невозможен?

Он убедился в этом, далеко уйдя по тому пути, на котором первые шаги делали еще Саккери и Ламберт. В качестве предположения он ввел утверждение, противоположное эвклидовскому постулату, а именно: «через точку, не лежащую на данной прямой, можно провести не одну, а по крайней мере две прямые, параллельные данной». Примем это утверждение условно как аксиому и, присоединив его к другим положениям геометрии, будем развивать отсюда дальнейшие следствия. Тогда, если такое утверждение несовместимо с другими положениями геометрии, мы придем к противоречию, и тем самым V постулат будет доказан от противного: противоположное предположение приводится к противоречию. Однако поскольку такого противоречия не обнаруживается, приходим к двум выводам, которые и сделал Лобачевский.

Первый вывод состоит в том, что V постулат не доказуем. Второй вывод состоит в том, что на основе противоположной, только что сформулированной аксиомы можно развивать цепь следствий — теорем, которые не будут заключать противоречия. Эти следствия образуют тем самым некоторую логически возможную, непротиворечивую теорию, которая может рассматриваться как новая, неэвклидова геометрия. Лобачевский осторожно назвал ее «воображаемой», поскольку не мог еще найти ее реального объяснения. Но логическая ее возможность для него была ясна. Высказывая и отстаивая это твердое убеждение, Лобачевский проявил истинное величие гения, который не колеблясь отстаивает свои убеждения, а не прячет их от общественного мнения, опасаясь непонимания и критики.

Итак, первые два вывода, полученные Лобачевским, состояли в утверждении недоказуемости V постулата и в возможности развить на основе

[1] Так писал Лобачевский в 1823 г. в своем курсе геометрии, который при его жизни не был, однако, опубликован. Этот курс «Геометрия» был впервые издан в 1910 г.

противоположной аксиомы новую геометрию, логически столь же содержательную и совершенную, как эвклидова, несмотря на противоречие ее выводов наглядным представлениям о пространстве. Лобачевский фактически развил эту новую геометрию, которая теперь носит его имя. В этом заключался общий результат огромной важности: *логически мыслима не одна геометрия.* Значение этого вывода во всем его объеме мы еще выясним; в нем, собственно, уже содержится не малая доля решения тех вопросов об абстрактных математических пространствах, которые были поставлены в начале этой главы.

Но вернемся к приведенному выше высказыванию Лобачевского. Он говорит, что геометрическую истину, подобно другим физическим законам, могут проверить лишь опыты. Это значит, во-первых, что под истиной нужно понимать соответствие отвлеченных понятий реальной действительности. Это соответствие может установить только опыт, и, следовательно, для проверки истинности тех или иных выводов нужны опытные исследования, одних же логических умозаключений для этого недостаточно. Хотя эвклидова геометрия отражает реальные свойства пространства очень точно, нельзя быть уверенным, что дальнейшие исследования не обнаружат только приближенную правильность эвклидовой геометрии как учения о свойствах реального пространства. Тогда геометрия как учение о реальном пространстве (а не как логическая система) потребует изменения и уточнения в согласии с новыми опытными данными.

Эта гениальная мысль Лобачевского нашла полное подтверждение в новом развитии физики — в теории относительности.

Сам Лобачевский предпринял вычисления на основе астрономических наблюдений, с тем чтобы проверить точность эвклидовой геометрии. Эти вычисления подтвердили тогда ее правильность в пределах доступной точности. Теперь положение изменилось, хотя нужно сразу оговорить, что и геометрия Лобачевского не оказалась более точной в применении к пространству, его свойства оказались другими, более сложными. Но еще раньше геометрия Лобачевского нашла свое обоснование и применение в другой связи, о чем мы будем подробно говорить дальше.

Нужно подчеркнуть, что Лобачевский вовсе не рассматривал свою геометрию просто как логическую схему, построенную на произвольно принятых предположениях. Главную задачу он видел не в логическом анализе оснований геометрии, а в исследовании их отношения к действительности. Поскольку опыт не может дать абсолютно точного решения вопроса о верности эвклидова постулата, постольку имеет смысл исследование тех логических возможностей, которые представляются более основными предпосылками геометрии. Это математическое исследование поможет наметить те пути, по которым должно идти физическое изучение свойств реального пространства. Тем более, что геометрия Эвклида есть предельный случай геометрии Лобачевского, и потому эта последняя включает более широкие возможности. С этой точки зрения ограничение

постулатом Эвклида было бы запретом для развития теории. Теория должна выходить за пределы уже известного, чтобы искать и указывать пути к открытию новых фактов и законов. Глубокое понимание связи математики с действительностью позволяет выделить из разнообразия логических возможностей именно те, которые имеют наибольшее основание оказаться полезными в познании природы. Если бы геометры вслед за Лобачевским не развивали математического учения о возможных свойствах пространства, современная физика не имела бы тех математических средств, которые позволили формулировать и развить положения теории относительности.

Итак, подведем итог того решения проблемы V постулата, которое дал Лобачевский.

1°. Постулат недоказуем.

2°. Присоединяя к основным положениям геометрии противоположную аксиому, можно развить логически совершенную и содержательную геометрию, отличную от эвклидовой.

3°. Правильность выводов той или иной логически мыслимой геометрии в применении к реальному пространству проверяется только опытом. Логически мыслимая геометрия должна разрабатываться не как произвольная логическая схема, а как теория, намечающая возможные пути и методы развития физических теорий.

Решение это совершенно отлично от того, что хотели получить геометры, пытавшиеся доказать эвклидов постулат. Оно настолько шло вразрез с установившимися представлениями, что не встретило понимания среди математиков. Оно было для них еще слишком новым и радикальным. Лобачевский как бы разрубил гордиев узел теории параллельных, а не распутал его, как то рассчитывали сделать другие.

2. Почти одновременно с Лобачевским недоказуемость V постулата и возможность неэвклидовой геометрии обнаружил венгерский геометр Янош Бóйаи (1802—1860), который опубликовал свои выводы в виде дополнения к вышедшему в 1832 г. геометрическому трактату своего отца Фаркаша Бойаи. Предварительно отец послал работу сына на отзыв Гауссу и получил похвальный ответ, где Гаусс сообщал, что уже давно сам пришел к тем же выводам. Однако от печатных выступлений Гаусс все же воздержался. В одном из писем он объясняет это тем, что боится быть непонятым.

В науке всегда бывает так, что назревшие в ней выводы почти одновременно и независимо получаются разными учеными. Интегральное и дифференциальное исчисление развивали одновременно Ньютон и Лейбниц; к идеям Дарвина независимо пришел в то же время Уоллес; начала теории относительности одновременно с Эйнштейном наметил также Пуанкаре, и таких примеров можно привести много. Они доказывают лишний раз, что наука развивается необходимым образом путем решения назревших в ней задач, а не путем случайных открытий и догадок. Так

и к открытию возможности неэвклидовой геометрии подошли одновременно несколько геометров: Лобачевский, Бо́йаи, Гаусс, Швейкарт и Тауринус.

Однако, как это также постоянно бывает в науке, не все ученые, пришедшие к новому результату, играют в его установлении одинаковую роль, и не всем можно приписать одинаковую заслугу. Здесь имеет значение и первенство по времени, и ясность и глубина выводов, и последовательность и обоснованность их проведения. Ни Швейкарт, ни Тауринус не были убеждены в правомерности новой геометрии, а это как раз и было решающим в данном случае, тем более что отдельные ее выводы были получены еще Саккери и Ламбертом. Гаусс, хотя и имел, повидимому, это убеждение, но не был настолько тверд в нем, чтобы рискнуть выступить с ним открыто. Глубокий и разносторонний математик, он не имел мужества отстаивать новые идеи и не произвел поэтому в математике такого переворота, какой сделал Лобачевский. В этом Гаусс оставался типичным представителем немецкой интеллигенции той эпохи, с ее глубокой теоретической мыслью и политической трусостью. Этот дух выразился в философии Гегеля — современника Гаусса, который, как говорит Ленин, «гениально угадал, именно угадал, не более», диалектику природы и познания, но, угадав эту, по словам Герцена, «алгебру революции», подчинил ее своей идеалистической, реакционной системе философии.

Бо́йаи не проявил нерешительности, но он не дал новым идеям столь далекого и глубоко идущего развития, какое дал Лобачевский. Именно Лобачевский первый открыто — устно в 1826 г. и печатно в 1829 г. — высказал новые идеи и продолжал развивать и пропагандировать их в ряде трудов, кончая вышедшей в 1855 г. «Пангеометрией», которую он диктовал, будучи уже слепым стариком на склоне лет, сохраняя твердость духа и уверенность в своей правоте. Именно поэтому новая геометрия по праву носит его имя.

Н. И. Лобачевский не только развил новую геометрию, но и правильно поставил вопрос об отношении геометрии к действительности. У идеалистов большим почетом пользовалась подновляемая до сих пор философия Канта, согласно которой пространство не является реальной формой существования материи, а лишь врожденной формой нашего представления, априорной, т. е. не зависящей от опыта, формой созерцания. По Канту, следовательно, и геометрия является априорной, не зависящей от опыта. Лобачевский опроверг это идеалистическое воззрение. Он выступил против него как материалист, дав глубокое материалистическое понимание отношения геометрии к действительности. Он говорил, что истину проверить могут только опыты. Вопреки Канту, Лобачевский утверждал, что «первые», т. е. основные «понятия приобретаются чувствами, врожденным — не должно верить». Лобачевский не только настаивал на происхождении представлений и понятий из опыта (о чем говорили материалисты

и до него), но и показал, что отношение геометрии к действительности должно еще уточняться опытом. Тут заключалась уже идея *развития* геометрии, идея неограниченного приближения к абсолютной истине по мере развития наших знаний.

Таким образом, Лобачевский был не только гениальным геометром, но и философом-материалистом. Он был также энергичным и разносторонним деятелем на поприще русского просвещения, состоя профессором и в течение почти 20 лет ректором Казанского университета. В материализме Лобачевского, в мужестве и стойкости, с которыми он отстаивал свои идеи, вопреки непониманию и даже издевательствам, в его широкой деятельности как ученого, педагога и организатора выразились характерные черты лучших представителей русского образованного общества того времени. То была эпоха бурного подъема и расцвета русского гения и общественного самосознания — эпоха Пушкина, декабристов, Белинского. Россия выступала на мировую арену не как робкая и послушная ученица европейской культуры, а как сила, дающая свое, новое, такое, чего не знала еще Европа. Так, в математике Лобачевский решил проблему, стоявшую перед наукой две тысячи лет, и повернул ее развитие на новые пути. Его материализм, его научное мужество сродни материализму и мужеству Радищева, Пестеля и Белинского. Лобачевский был не только создателем новой геометрии, но ученым, мыслителем и гражданином — великим человеком в полном смысле этого слова. Последовавшее за Лобачевским развитие геометрии, конечно, далеко вышло за пределы, о каких он мог мыслить, но по праву от него нужно считать начало новой эпохи в геометрии и в математике вообще.

§ 3. ГЕОМЕТРИЯ ЛОБАЧЕВСКОГО

1. Итак, Лобачевский принял за основу утверждение, противоположное V постулату: в данной плоскости через точку можно провести по крайней мере две прямые, не пересекающие данную прямую. Отсюда он вывел ряд далеко идущих следствий, которые и образовали новую геометрию. Эта геометрия строилась, следовательно, как некоторая мыслимая теория, как совокупность теорем, логически доказываемых: исходя из сделанного предположения, в соединении с другими [1] основными посылками эвклидовой или, как говорил Лобачевский, «употребительной» геометрии.

В своих выводах Лобачевский получил все результаты, аналогичные результатам «употребительной» элементарной геометрии, т. е. дошел до неэвклидовой тригонометрии и решения треугольников, до вычисления площадей и объемов. Мы не можем здесь проследить эту цепь выводов Лобачевского не потому, что они слишком сложны, а прежде всего по недостатку места. Ведь и школьный курс «употребительной» геометрии

[1] Эти, так сказать, «остальные» положения геометрии будут ниже (в § 5) тоже точно сформулированы.

довольно велик, а выводы Лобачевского, конечно, не проще и не короче этих «употребительных» выводов. Поэтому мы отметим здесь только некоторые поразительные результаты Лобачевского, отсылая читателя, интересующегося более глубоким изучением неэвклидовой геометрии, к специальной литературе. Дальше же мы выясним простой реальный смысл неэвклидовой геометрии.

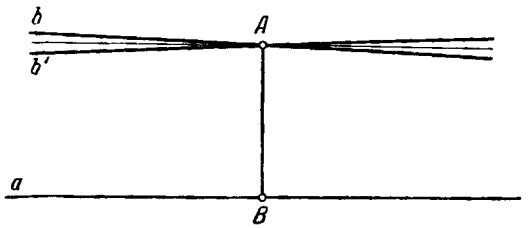


Рис. 2.

Начнем с теории параллельных линий. Пусть дана прямая a и точка A вне ее. Опустим из A перпендикуляр AB на прямую a. По основному предположению существуют по крайней мере две прямые, проходящие через точку A и не пересекающие нашу прямую a. Тогда всякая прямая, идущая в угле между этими прямыми, также не пересекает a. На рис. 2 прямые b и b' при продолжении пересекут a вопреки предположению Лобачевского. Но в этом нет ничего удивительного. Ведь Лобачевский рассуждал не о чертежах, которые мы делаем на обычной плоскости; он развивал логические следствия из своего предположения, которое

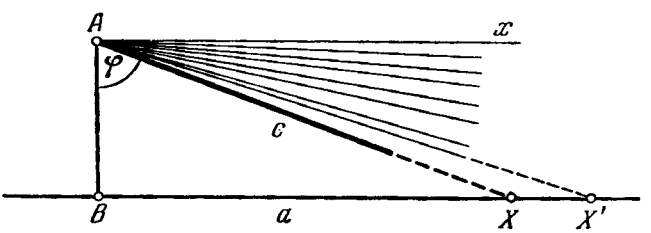


Рис. 3.

противоречит тому, что мы привыкли видеть на чертежах. Чертежи здесь играют только вспомогательную роль; на них не изображаются в точности факты неэвклидовой геометрии, поскольку мы проводим на чертеже обычные, в пределах точности чертежа — безусловно эвклидовы — прямые на обычной плоскости.

Это противоречие между логической возможностью и наглядным представлением было главной трудностью в понимании геометрии Лобачевского. Но если речь идет о геометрии как логической теории, то нужно заботиться о логической строгости рассуждений, а не о согласии с привычными чертежами.

2. Вернемся опять к нашей прямой a и точке A. Проведем из A полупрямую x, не пересекающую a (например, перпендикулярную AB), и будем вращать ее вокруг A так, чтобы угол φ между AB и x уменьшался, не доводя, однако, эту полупрямую до пересечения с a. Тогда полупря-

мая x будет стремиться к предельному положению, отвечающему наименьшему значению угла φ. Эта предельная полупрямая c также не будет пересекать a.

Действительно, если бы она пересекала прямую a в какой-то точке X (рис. 3), то мы могли бы взять точку X' правее и получили бы полупрямую AX', пересекающую a, но образующую с AB больший угол. А это невозможно, так как по построению полупрямой c любая полупрямая x, образующая с AB бóльший угол, прямую a уже не пересекает.

Следовательно, полупрямая c не пересекает a и является вместе с тем крайней из всех полупрямых, проходящих через точку A и не пересекающих прямую a.

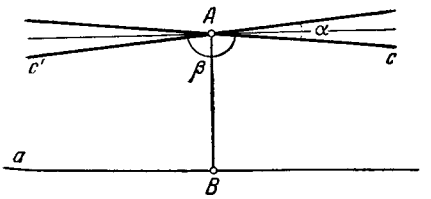


Рис. 4.

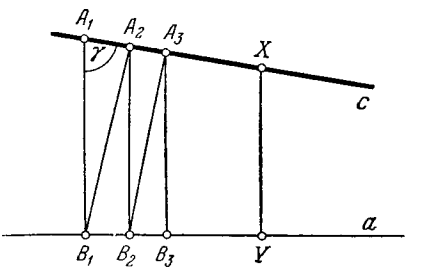


Рис. 5.

По симметрии очевидно, что с другой стороны также можно провести полупрямую c', не пересекающую a и самую крайнюю из всех таких полупрямых. Если бы c и c' служили продолжением друг друга, то они образовали бы вместе одну прямую $c+c'$. Эта прямая была бы тогда единственной прямой, параллельной a и проходящей через данную точку A, потому что при малейшем ее вращении или c, или c' должна была бы пересечь a. А раз предположено, что параллельная не одна, а их по крайней мере две, то полупрямые c и c' не являются продолжением друг друга.

Итак, мы доказали первую теорему геометрии Лобачевского:

Из точки A, не лежащей на данной прямой a, можно провести две полупрямые c и c' так, что они не пересекают прямую a, но любая полупрямая, идущая в угле между ними, пересекает прямую a.

Если продолжить полупрямые c и c', то получим (рис. 4) две прямые, не пересекающие a и обладающие тем свойством, что всякая прямая, проходящая через A в угле α между этими прямыми, не пересекает прямую a, а всякая прямая, проходящая в угле β, пересекает прямую a. Такие прямые c, c' Лобачевский назвал *параллельными* прямой a: прямую c — параллельной справа, прямую c' — параллельной слева. Половину угла β Лобачевский назвал *углом параллельности*; он меньше прямого, поскольку β меньше двух прямых.

3. Рассмотрим теперь, как меняется расстояние от точки X на прямой c до прямой a при движении X вдоль c (рис. 5). В эвклидовой

геометрии расстояние между параллельными прямыми постоянно. Здесь же мы убедимся, что при движении точки X направо ее расстояние от a (т. е. длина перпендикуляра XY) убывает.

Опустим из точки A_1 перпендикуляр A_1B_1 на прямую a. Из точки B_1 опустим перпендикуляр B_1A_2 на прямую c (точка A_2 лежит правее A_1, так как угол γ острый). Наконец, из точки A_2 опустим перпендикуляр A_2B_2 опять на прямую a. Докажем, что A_2B_2 меньше A_1B_1.

Теорема о том, что перпендикуляр короче наклонной, верна и в геометрии Лобачевского, поскольку ее доказательство (которое можно найти в любом школьном учебнике геометрии) не опирается на понятие о параллельных прямых или связанные с ними выводы. А раз перпендикуляр короче наклонной, то B_1A_2 как перпендикуляр к прямой c короче A_1B_1 и аналогично A_2B_2 как перпендикуляр к a короче B_1A_2. Следовательно, A_2B_2 короче A_1B_1.

Далее, опуская на прямую c перпендикуляр B_2A_3 из точки B_2 и повторяя то же рассуждение, можно убедиться, что A_3B_3 короче A_2B_2. Продолжая это построение, мы получим последовательность все более и более коротких перпендикуляров, т. е. расстояния точек A_1, A_2, ... от прямой a убывают. Дальше, дополняя наше простое рассуждение, можно было бы доказать, что вообще если точка X'' на c лежит правее X', то перпендикуляр $X''Y''$ короче $X'Y'$. Мы не будем на этом останавливаться. Проведенное рассуждение, мы надеемся, достаточно выясняет суть дела, а строгие доказательства не входят в нашу задачу.

Но замечательным оказывается, что, как можно доказать, расстояние XY не только убывает при движении точки X по прямой c вправо, но оно стремится к нулю, когда точка X удаляется в бесконечность. То-есть *параллельные прямые a и c асимптотически сближаются!* Вместе с тем можно доказать, что в противоположном направлении расстояние между ними не только увеличивается, но и растет до бесконечности.

В эвклидовой геометрии прямая, параллельная данной, проходит от нее на постоянном расстоянии. В геометрии же Лобачевского вообще не существует таких пар прямых, там прямые всегда расходятся до бесконечности или в одну сторону или в обе стороны. Линия же, проходящая на постоянном расстоянии от данной прямой, никогда не будет прямой, а является некоторой кривой, называемой *эквидистантой*, т. е. равноудаленной.

Эти выводы геометрии Лобачевского поистине поразительны и никак не вяжутся с привычными наглядными представлениями. Но как мы уже говорили, такое несоответствие не может быть аргументом против геометрии Лобачевского как отвлеченной теории, логически развиваемой из принятых предпосылок.

4. Рассмотрим теперь еще угол параллельности, т. е. тот угол γ, который образует прямая c, параллельная данной прямой a, с перпендикуляром CA (рис. 6). Докажем, что этот угол тем меньше, чем дальше

точка C от прямой a. Для этого докажем сначала следующее. Если две прямые b и b' образуют с секущей BB' равные углы α, α', то эти прямые имеют общий перпендикуляр (рис. 7).

Для доказательства проведем через середину O отрезка BB' прямую CC', перпендикулярную прямой b. Получим два треугольника OBC и $OB'C'$. Их стороны OB и OB' равны по построению. Углы при общей вершине O равны, как вертикальные. Угол α'' равен углу α', так как они тоже вертикальные. Угол же α равен углу α' по условию. Следовательно, угол α равен также углу α''. Таким образом, у наших треугольников OBC и $OB'C'$ равны стороны OB и OB' и прилежащие к ним углы. А тогда по известной теореме треугольники равны, т. е. равны, в частности, их углы при C и C'. Но угол C прямой, так как по построению прямая CC' перпендикулярна b. Следовательно, угол C' также прямой, т. е. CC' перпендикулярна также и к прямой b'. Таким образом, отрезок CC' является общим перпендикуляром к обеим прямым b и b'. Существование общего перпендикуляра доказано.

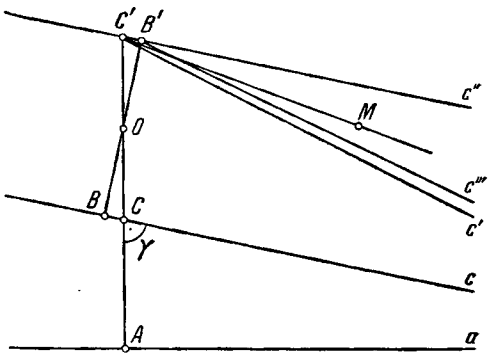


Рис. 6.

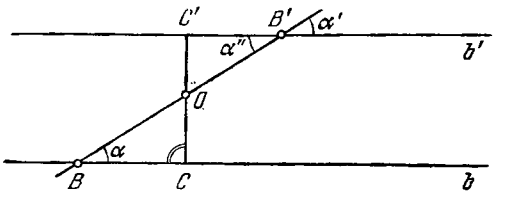


Рис. 7.

Теперь докажем, что угол параллельности убывает с увеличением расстояния от прямой. То-есть, если точка C' лежит от прямой a дальше, чем C, как на рис. 6, то параллель c', идущая из C', образует с перпендикуляром $C'A$ меньший угол, чем параллель c, идущая из C.

Для доказательства проведем из C' прямую c'' под тем же углом к $C'A$, под каким идет параллель c. Тогда прямые c и c'' образуют с секущей CC' равные углы. Поэтому, как только что доказано, они имеют общий перпендикуляр BB'. Тогда из основания B' этого перпендикуляра можно провести прямую c''', параллельную c и образующую с перпендикуляром угол меньше прямого, потому что, как мы уже знаем, параллель образует с перпендикуляром угол меньше прямого. Возьмем теперь в угле между c'' и c''' любую точку M и проведем прямую $C'M$. Она войдет в угол между c'', c''' и дальше уже не сможет пересечь c'''. Тем более она не будет пересекать прямую c. Но она образует с AC' уже меньший угол, чем c'', т. е. угол меньший γ. Тем более, еще меньший угол будет образовывать параллель c', поскольку она является самой крайней

из всех прямых, идущих из C' и не пересекающих a. Следовательно, параллель c' образует с $C'A$ угол меньший, чем c, а это и значит, что угол параллельности убывает при переходе к более далекой точке C', что и требовалось доказать.

Итак, мы доказали, что угол параллельности убывает по мере удаления точки C от прямой a. Но оказывается, можно доказать и более того: *если точку C удалять в бесконечность, то этот угол стремится к нулю.* То-есть на сколь угодно большом расстоянии от прямой a параллель к ней

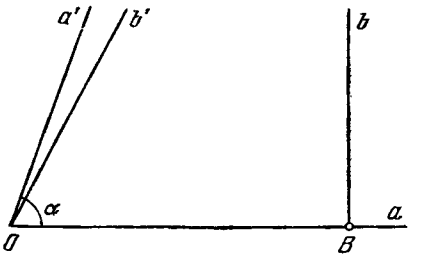


Рис. 8.

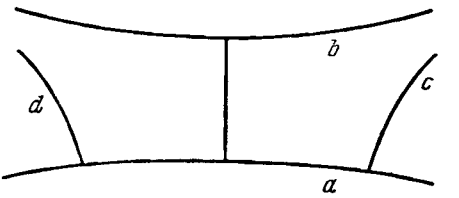


Рис. 9.

образует с перпендикуляром к этой прямой сколь угодно малый угол [1]. Иначе говоря, если очень далеко от прямой a отклониться от перпендикуляра к ней на очень малый угол, то мы вовсе не пересечем прямую a, идя по «отклонившейся» прямой. Этот факт геометрии Лобачевского также производит удивительное впечатление. Но дальше можно получить и другие, не менее удивительные результаты.

Например, возьмем острый угол α, образуемый полупрямыми a и a'. Проведем перпендикуляр b к a настолько далеко от вершины O угла α, чтобы угол параллельности, соответствующий взятому расстоянию OB (рис. 8), был меньше α. Раз угол α больше угла параллельности, то прямая b', проведенная из O параллельно b, образует с a меньший угол. Но она не пересекает прямую b. Следовательно, a' тем более ее не пересекает. Этим доказано, что перпендикуляр к стороне острого угла, проведенный достаточно далеко от вершины, не пересекает другую сторону.

5. Мы привели все предыдущие выводы с двоякой целью. Во-первых, и это главное, мы хотели показать на самых простых примерах, как фактически можно получать теоремы геометрии Лобачевского, исходя из принятых предпосылок. Это служит простейшим примером того, как вообще математики получают выводы в абстрактной геометрии, как вообще можно получить такие выводы, не связанные с привычными наглядными

[1] Если h — длина перпендикуляра, γ — угол параллельности, то, как показал Лобачевский, $\operatorname{tg}\frac{\gamma}{2} = e^{-\frac{h}{k}}$, где k — постоянная, зависящая от единиц длины, а e — известное основание натуральных логарифмов. Очевидно $e^{-\frac{h}{k}}$, а с ним и γ стремится к нулю при $h \to \infty$.

представлениями. Во-вторых, мы хотели показать, сколь своеобразные результаты получаются в геометрии Лобачевского. Приведем еще примеры.

Две прямые в плоскости Лобачевского либо пересекаются, либо параллельны в смысле Лобачевского, и тогда они в одну сторону асимптотически сближаются, а в другую — бесконечно расходятся, либо они имеют общий перпендикуляр и в обе стороны от него бесконечно расходятся.

Если прямые a, b имеют общий перпендикуляр (рис. 9), то к прямой a можно провести два перпендикуляра c, d, параллельные (в смысле Лобачевского) прямой b, вся прямая b лежит в полосе между прямыми c, d.

Предел окружностей бесконечно увеличивающегося радиуса не есть прямая, а некоторая кривая, называемая *предельной окружностью*. Через три точки, не лежащие на одной прямой, не всегда можно провести окружность, а либо окружность, либо предельную окружность, либо эквидистанту (т. е. линию, образованную точками, равноудаленными от некоторой прямой).

Сумма углов треугольника всегда меньше двух прямых. Если треугольник увеличивается так, что все три его высоты неограниченно возрастают, то все три его угла стремятся к нулю.

Не существует треугольников сколь угодно большой площади.

Два треугольника равны, если их углы равны.

Длина окружности l не пропорциональна радиусу r, а растет быстрее (в основном по показательному закону). Именно, имеет место формула

$$l = \pi k\left(e^{\frac{r}{k}} - e^{-\frac{r}{k}}\right), \tag{1}$$

где k — постоянная, зависящая от единиц длины. Так как

$$e^{\frac{r}{k}} = 1 + \frac{r}{k} + \frac{1}{2}\left(\frac{r}{k}\right)^2 + \ldots, \qquad e^{-\frac{r}{k}} = 1 - \frac{r}{k} + \frac{1}{2}\left(\frac{r}{k}\right)^2 - \ldots,$$

то из формулы (1) получаем:

$$l = 2\pi r\left(1 + \frac{1}{6}\frac{r^2}{k^2} + \ldots\right). \tag{2}$$

И только при малых отношениях $\frac{r}{k}$ с достаточной точностью получается, что $l = 2\pi r$.

Все эти выводы являются логическими следствиями принятых предпосылок: «аксиомы Лобачевского» в соединении с основными положениями «употребительной» геометрии.

6. Чрезвычайно важное свойство геометрии Лобачевского заключается в том, что в достаточно малых областях она мало отличается от геометрии Эвклида; чем меньше область, тем это различие меньше.

Так, для достаточно малых треугольников связь сторон и углов с достаточной точностью выражается формулами обычной тригонометрии и притом тем точнее, чем меньше треугольник.

Формула (2) показывает, что при малых радиусах длина окружности с хорошей точностью пропорциональна радиусу. Точно так же сумма углов треугольника мало отличается от двух прямых и т. п.

В формулу для длины окружности входит постоянная k, зависящая от единиц длины. Если радиус мал в сравнении с k, т. е. если $\frac{r}{k}$ мало, то, как видно из формулы (2), длина l близка к $2\pi r$. Вообще, чем меньше отношение размеров фигуры к этой постоянной, тем точнее свойства фигуры подходят к свойствам соответствующей фигуры в эвклидовой геометрии[1].

Мерой отклонения свойств фигуры геометрии Лобачевского от свойств фигуры эвклидовой геометрии служит отношение $\frac{r}{k}$, если r измеряет размеры фигуры (радиус окружности, стороны треугольника и т. п.).

Отсюда вытекает важный вывод.

Пусть мы имеем дело с реальным пространством и измеряем расстояние в километрах. Допустим, что постоянная k при этом очень велика, скажем равна 10^{12}.

Тогда, например, по формуле (2) для окружности с радиусом даже в 100 км отношение ее длины к радиусу будет отличаться от 2π меньше, чем на одну миллиардную. Того же порядка будут отклонения от других соотношений эвклидовой геометрии. В пределах одного километра они будут уже порядка $\frac{1}{k}$, т. е. 10^{-12}, а в пределах метра — порядка 10^{-15}, т. е. будут совсем ничтожными. Таких отклонений от эвклидовой геометрии уже нельзя было бы заметить, потому что даже размеры атома в сто раз больше (они составляют величину порядка 10^{-13} км).

---

[1] Например, если a, b, c — катеты и гипотенуза прямоугольного треугольника, то вместо теоремы Пифагора имеет место соотношение

$$2\left(e^{\frac{c}{k}}+e^{-\frac{c}{k}}\right)=\left(e^{\frac{a}{k}}+e^{-\frac{a}{k}}\right)\left(e^{\frac{b}{k}}+e^{-\frac{b}{k}}\right).$$

Разлагая в ряды, получим: $c^2+\frac{c^4}{12k^2}+\ldots=a^2+b^2+\frac{a^4+6a^2b^2+b^4}{12k^2}+\ldots$, так что при большом k получаем теорему Пифагора $c^2=a^2+b^2$. Далее, по формуле Лобачевского для угла параллельности γ (см. сноску на стр. 106) $\operatorname{tg}\frac{\gamma}{2}=e^{-\frac{h}{k}}$. Если $\frac{h}{k}$ мало, т. е. если параллельные близки, то $\operatorname{tg}\frac{\gamma}{2}=e^{-\frac{h}{k}}\approx 1$ и $\gamma\approx 90°$. Таким образом, на малых расстояниях параллельные на плоскости Лобачевского мало отличаются от эвклидовых.

С другой стороны, для астрономических масштабов отношение $\frac{r}{k}$ могло бы оказаться уже не слишком малым.

Поэтому Лобачевский и допускал, что, хотя в обычных масштабах геометрия Эвклида и верна с большой точностью, отклонения от нее можно будет заметить посредством астрономических наблюдений. Как уже было сказано, само это допущение оправдалось, но те незначительные отклонения от эвклидовой геометрии, которые обнаружены теперь в астрономических масштабах, оказываются еще более сложными.

Наконец, из приведенных рассуждений следует еще другой важный вывод. Именно: раз отклонения от эвклидовой геометрии тем меньше, чем больше постоянная k, то в пределе, при неограниченном увеличении ее, геометрия Лобачевского переходит в геометрию Эвклида. То-есть *геометрия Эвклида есть не более как предельный случай геометрии Лобачевского.* Поэтому если к геометрии Лобачевского присоединить и этот ее предельный случай, то она охватит также геометрию Эвклида, оказываясь в этом смысле более общей теорией. Соответственно этому Лобачевский называл свою теорию «пангеометрией», т. е. общей геометрией. Такое соотношение теорий постоянно обнаруживается в развитии математики и естествознания: новая теория включает старую как предельный случай, соответственно движению познания от более частных выводов к более общим.

Однако все приведенные рассуждения и выводы оставались как бы мало понятной игрой ума, если бы не был установлен сравнительно простой реальный смысл геометрии Лобачевского в системе уже привычных понятий эвклидовой геометрии. Решение этой задачи не было доведено до конца самим Лобачевским; оно осталось на долю его последователей и было получено почти 40 лет спустя после выхода в свет первой его работы. В чем состоит это решение, мы расскажем в следующем параграфе.

§ 4. РЕАЛЬНЫЙ СМЫСЛ ГЕОМЕТРИИ ЛОБАЧЕВСКОГО

1. Наглядное истолкование геометрии Лобачевского было получено впервые в 1868 г., когда итальянский геометр Бельтрами заметил, что внутренняя геометрия на некоторой поверхности — псевдосфере — совпадает с геометрией на куске плоскости Лобачевского. Напомним, что под внутренней геометрией поверхности понимают ту совокупность свойств фигур на ней, которые определяются только лишь измерением длин на самой поверхности. На рис. 10 слева изображена так называемая *трактриса.* Это кривая, обладающая тем свойством, что длина отрезка ее касательной от точки касания до пересечения с осью Oy для всех точек кривой постоянна. Ось Oy ее асимптота. Вращая трактрису

вокруг ее асимптоты, мы получаем изображенную на рис. 10 справа поверхность, которая и называется *псевдосферой*.

Истолкование геометрии Лобачевского по Бельтрами сводится к тому, что все геометрические соотношения на куске плоскости Лобачевского совпадают с геометрическими соотношениями на подходящем куске псевдосферы, если принять следующие условия. Роль прямолинейных отрезков играют кратчайшие линии на поверхности — геодезические. Расстояние между точками определяется как длина наиболее

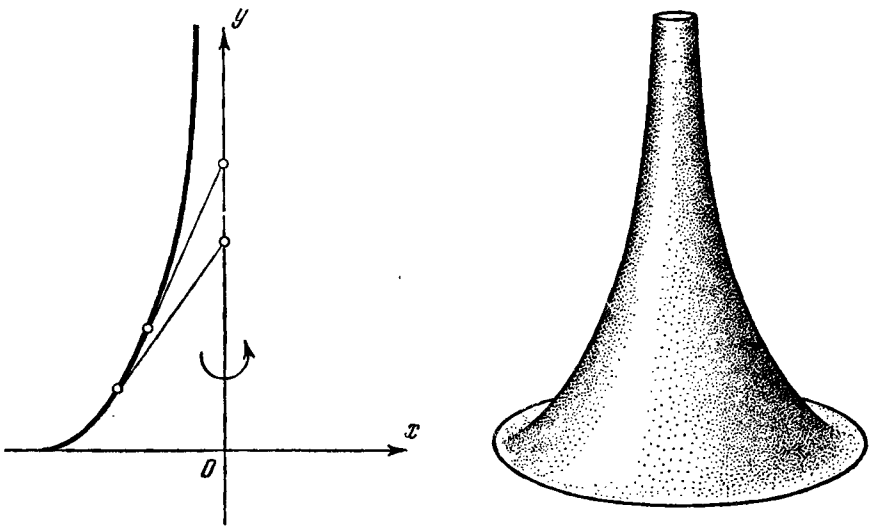


Рис. 10.

короткой линии, соединяющей их на поверхности. Фигуры считаются равными, если можно так сопоставить их точки, что внутренние расстояния между соответственными точками будут равны. Перемещение фигур на псевдосфере, сохраняющее их размеры с точки зрения внутренней геометрии, хотя и сопровождается изгибанием, изображает движение в плоскости Лобачевского. Длины, углы и площади измеряются на поверхности, как обычно и отвечают длинам, углам и площадям в геометрии Лобачевского.

Истолкование Бельтрами показывает, что при этих условиях, каждому утверждению геометрии Лобачевского, относящемуся к куску плоскости, отвечает непосредственный факт внутренней геометрии псевдосферы. Геометрия Лобачевского имеет стало быть совершенно реальный смысл: она есть не что иное, как абстрактно изложенная геометрия на псевдосфере.

Нужно сказать, что внутреннюю геометрию псевдосферы за 30 лет до открытия Бельтрами уже исследовал Ф. Миндинг, который фактически установил свойства, показывающие ее совпадение с геометрией

Лобачевского. Однако этого ни он, ни кто-либо другой не заметил, пока идеи Лобачевского не получили достаточного распространения. Бельтрами оставалось только сопоставить выводы Лобачевского и Миндинга, чтобы увидеть их связь.

Открытие Бельтрами сразу изменило отношение математиков к геометрии Лобачевского; из «воображаемой» она стала реальной[1].

2. Однако, как было подчеркнуто, на псевдосфере реализуется геометрия не всей плоскости Лобачевского, а лишь ее куска[2]. Поэтому оставалась еще не решенной задача выяснения реального смысла геометрии Лобачевского на всей плоскости и тем более в пространстве. Эта задача и была вскоре, в 1870 г., разрешена немецким математиком Клейном. Изложим, в чём состояло данное им решение.

Возьмем на обычной эвклидовой плоскости круг и будем рассматривать лишь внутренность этого круга, т. е. исключим из рассмотрения его окружность и область вне круга. Эту внутренность круга назовем условно «плоскостью» — она, оказывается, и будет играть роль плоскости Лобачевского. Хорды нашего круга назовем «прямыми», причем согласно принятому условию концы хорд как лежащие на окружности исключаются. Наконец, назовем «движением» любое такое преобразование круга, которое переводит его самого в себя и оставляет прямые прямыми, т. е. не искривляет его хорд. Простейшим примером

---

[1] История установления реального смысла геометрии Лобачевского была в действительности еще более сложной. Во-первых, уже сам Лобачевский владел средствами доказательства ее непротиворечивости посредством так называемой аналитической модели, но не мог еще до конца выразить такое доказательство. Это было сделано гораздо позже. Во-вторых, немецкий математик Риман в 1854 г. выступил с теорией (см. § 10), в которой уже содержались выводы Бельтрами, но Риман не выразил их явно; его доклад не был понят и был опубликован лишь после его смерти в том же 1868 г., когда появилась работа Бельтрами. Вообще вся история геометрии Лобачевского от попыток доказательства постулата Эвклида до полного выяснения значения неэвклидовой геометрии чрезвычайно поучительна тем, что показывает, каких усилий и обходных путей требует часто открытие истины, которая потом оказывается простой и понятной.

[2] Псевдосфера имеет всюду одинаковую отрицательную гауссову кривизну. Все поверхности постоянной отрицательной кривизны имеют (по крайней мере в малых кусках) ту же внутреннюю геометрию и тем самым могут служить для изображения геометрии Лобачевского. Однако, как доказал в 1901 г. Гильберт, никакая из этих поверхностей не может быть бесконечно продолжена во все стороны без особенностей и потому не может реализовать *всю* плоскость Лобачевского. С другой стороны, в 1955 г. молодой голландский математик Кейпер установил, что существуют гладкие поверхности, реализующие в смысле их внутренней геометрии всю плоскость Лобачевского, но такие поверхности, хотя и гладкие, не могут быть непрерывно изогнутыми, они не имеют определенной кривизны.

Заметим еще следующее. При реализации геометрии Лобачевского на поверхности постоянной отрицательной кривизны K постоянная k, фигурировавшая в формулах предыдущего параграфа, получает простой смысл: $k^2 = -\frac{1}{K}$.

такого преобразования служит вращение круга вокруг центра, но оказывается, что таких преобразований гораздо больше. Каковы эти преобразования, будет сказано дальше.

Если ввести такие условные обозначения, то оказывается, что факты обычной геометрии внутри нашего круга превращаются в теоремы геометрии Лобачевского. И обратно: всякая теорема геометрии Лобачевского истолковывается как факт обычной геометрии внутри круга.

Например, по аксиоме Лобачевского через точку, не лежащую на данной прямой, можно провести по крайней мере две прямые, не пересекающие данную. Переводим эту аксиому на язык обычной геометрии согласно принятому условию, т. е. заменяя прямые хордами. Тогда мы получим утверждение: через точку внутри круга, не лежащую на данной хорде, можно провести по крайней мере две хорды, не пересекающие данную. Справедливость этого утверждения очевидна из рис. 11. Следовательно, аксиома Лобачевского здесь выполняется.

Вспомним далее, что в геометрии Лобачевского среди прямых, проходящих, через данную точку и не пересекающих данную прямую, есть две крайние, — те, которые по Лобачевскому только и называются параллельными данной прямой. Это означает, что среди хорд, проходящих через данную точку A и не пересекающих данную хорду BC, есть две крайние хорды. И действительно, такими крайними хордами будут хорды, подходящие одна к точке B, другая — к точке C. Они ведь не имеют с хордой BC общих точек, поскольку точки, лежащие на окружности, мы исключаем. Таким образом, эта теорема Лобачевского здесь выполняется.

Для дальнейшего перевода теорем Лобачевского на язык обычной геометрии внутри круга необходимо выяснить, как должны измеряться в круге отрезки и углы, чтобы это измерение отвечало геометрии Лобачевского. Конечно, это измерение не будет совпадать с обычным, потому что в обычном смысле хорда имеет конечную длину, а прямая, которую изображает хорда, бесконечна. В этом можно было бы даже усмотреть некоторое противоречие, но мы увидим, что никакого противоречия тут нет.

Прежде всего вспомним, что измерение длин отрезков производится следующим образом. Выбирается какой-либо отрезок AB, длина которого принимается равной единице, и длина любого другого отрезка XY определяется сравнением его с отрезком AB. При этом отрезок AB откладывается вдоль отрезка XY. Если остается еще доля отрезка XY, меньшая AB, то отрезок AB делят, например, на 10 равных частей (равных в том смысле, что каждая получается из другой движением); эти доли откладывают на оставшейся части отрезка XY; потом, если нужно, делят отрезок AB на 100 частей и т. д. В результате длина отрезка XY выразится в виде десятичной дроби, которая может быть и бесконечной. Стало быть, измерение длины производится путем пере-

движения целого или части отрезка, принятого за единицу, т. е. измерение основано на движении. А раз движения уже определены (мы определили их в данном случае как преобразования круга, переводящие прямые в прямые), то тем самым известно, какие отрезки считаются равными и как нужно измерять длину. Словом, определение движения уже включает, хотя и в неявном виде, закон измерения длин. Совершенно так же углы измеряют откладыванием угла, принятого за единицу. Таким образом, закон измерения углов также содержится в определении движения.

Законы измерения длин и углов, отвечающие геометрии Лобачевского, получаются довольно простыми, хотя и существенно отличными от

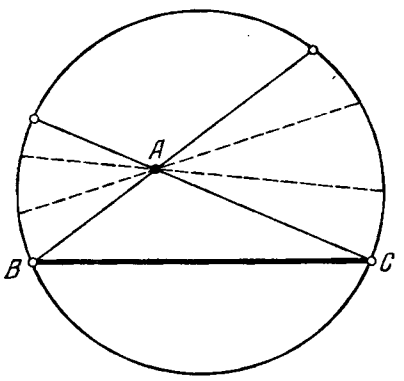


Рис. 11.

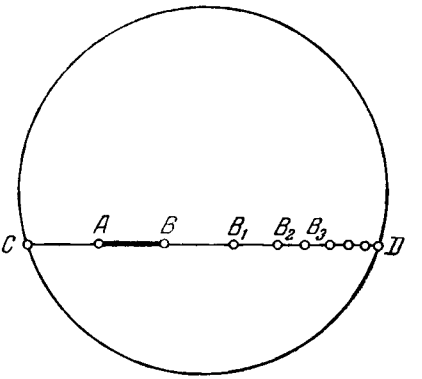


Рис. 12.

обычных. Приводить их вывод мы не будем, так как это не имеет в наших рассуждениях принципиального значения[1].

Закон измерения длин оказывается таким, что хорда имеет бесконечную длину. И это потому, что если путем преобразования, принятого нами за движение, перевести отрезок AB в отрезок BB_1, далее в отрезок B_1B_2 и т. д., то получаемые отрезки B_kB_{k+1} будут становиться в обычном смысле все короче (хотя и будут равными в смысле нашей модели геометрии Лобачевского; рис. 12). Точки B_1, B_2, ..., B_k, ... будут сгущаться к концу хорды. Но хорда у нас не имеет конца: конец ее по условию исключен, и она уже в этом смысле «бесконечна». В смысле геометрии Лобачевского точки B_1, B_2, ... никуда не будут сгущаться, они будут уходить в бесконечность. Путем преобразования, принятого нами за движение, откладывая друг за другом равные отрезки, нельзя изнутри круга достигнуть его окружности.

[1] Закон измерения длин оказывается следующим. Пусть отрезок AB лежит на хорде CD (рис. 12). Измеряем отрезки обычным образом и составляем так называемое сложное отношение $\frac{CB}{CA} : \frac{DB}{DA}$. Его логарифм принимают за длину отрезка AB.

Для того чтобы лучше понять, как в модели откладывают отрезок, рассмотрим то преобразование, которое играет роль переноса вдоль прямой.

Пусть на плоскости введены прямоугольные координаты с началом в центре круга. Для определенности будем считать, что наш круг имеет радиус, равный единице, так что его окружность представляется уравнением $x^2+y^2=1$, а точки внутри круга удовлетворяют неравенству $x^2+y^2<1$.

Рассмотрим преобразование, заданное формулами

$$x'=\frac{x+a}{1+ax}, \quad y'=\frac{y\sqrt{1-a^2}}{1+ax}, \tag{3}$$

где x', y' — кординаты точки, в которую переходит после преобразования точка, первоначально имевшая координаты x, y, а a — любое данное число, по абсолютной величине меньше единицы.

Если из формул (3) найти x и y, то получим, как легко проверить, обратное выражение x, y через x', y'

$$x=\frac{x'-a}{1-ax'}, \quad y=\frac{y'\sqrt{1-a^2}}{1-ax'}. \tag{4}$$

Преобразование (3) удовлетворяет двум условиям для «движений» в нашей модели: 1) оно переводит круг сам в себя; 2) оно переводит прямые в прямые.

Для доказательства первого свойства надо, собственно говоря, убедиться, что неравенство или равенство $x^2+y^2\leqslant 1$ влечет за собой соответствующее соотношение $x'^2+y'^2\leqslant 1$ и обратно. Покажем, например, что при $x^2+y^2=1$ обязательно и $x'^2+y'^2=1$, т. е. что точки, лежавшие на окружности данного круга, остаются снова на ней.

Вычисляем $x'^2+y'^2$, пользуясь формулами (3) и считая $x^2+y^2=1$, т. е. $y^2=1-x^2$,

$$x'^2+y'^2=\frac{(x+a)^2+y^2(1-a^2)}{(1+ax)^2}=\frac{(x+a)^2+(1-x^2)(1-a^2)}{(1+ax)^2}=$$

$$=\frac{x^2+2ax+a^2+1-x^2-a^2+a^2x^2}{(1+ax)^2}=\frac{1+2ax+a^2x^2}{1+2ax+a^2x^2}=1.$$

Следовательно, при $x^2+y^2=1$ также $x'^2+y'^2=1$. Аналогично проверяются остальные случаи.

Второе свойство преобразования (3) устанавливается столь же просто. В самом деле, мы знаем, что всякая прямая представляется линейным уравнением, и, обратно, всякое линейное уравнение представляет прямую. Пусть дана прямая

$$Ax+By+C=0. \tag{5}$$

После преобразования (4) получим

$$A\frac{x'-a}{1-ax'}+B\frac{y'\sqrt{1-a^2}}{1-ax'}+C=0,$$

или, приводя к общему знаменателю,

$$(A-aC)x'+B\sqrt{1-a^2}\,y'+(C-aA)=0.$$

Это уравнение линейное и, стало быть, представляет прямую. Это и есть та прямая, в которую при преобразовании переходит прямая (5).

Отметим еще, что преобразование (3) переводит ось Ox саму в себя, вызывая лишь смещение точек вдоль нее. Это ясно, так как на этой оси $y=0$, а по формуле (3) тогда также $y'=0$. На оси Ox преобразование задается одной формулой

$$x'=\frac{x+a}{1+ax} \quad (|a|<1). \tag{3'}$$

На этой прямой отрезок x_1x_2 переходит в отрезок $x_1'x_2'$ по формуле (3'), и по условию эти отрезки считаются равными. Так и происходит «откладывание отрезка».

Для центра O круга $x=0$ и, соответственно, $x'=a$, т. е. при преобразовании (3') центр переходит в точку A с координатой $x=a$.

Так как a можно задать любое, лишь бы было $|a|<1$, то центр можно перевести в любую точку на диаметре вдоль оси Ox.

При том же преобразовании точка, ранее находившаяся в A, перейдет в точку A_1 с координатой

$$x_1=\frac{a+a}{1+a^2}=\frac{2a}{1+a^2}.$$

Таким образом, отрезок OA при преобразовании (3) переходит в отрезок AA_1 — происходит «откладывание» этого отрезка на «прямой», изображаемой диаметром круга.

Повторяя то же преобразование, мы можем отложить тот же отрезок дальше сколько угодно раз. Точка A_n с координатой x_n будет переходить в точку A_{n+1} с координатой

$$x_{n+1}=\frac{x_n+a}{1+ax_n}.$$

Так мы будем получать точки A, A_1, A_2, ... с координатами

$$x_0=a,\quad x_1=\frac{2a}{1+a^2},\quad x_2=\frac{x_1+a}{1+ax_1}=\frac{3a+a^3}{1+3a^2},\ \ldots$$

Поскольку все отрезки A_nA_{n+1} получаются из OA преобразованием, изображающим движение, все они «равны» между собою — равны в смысле геометрии Лобачевского как она изображается на модели.

Легко доказать, что точки A_n сгущаются к концу диаметра. В смысле модели они удаляются в бесконечность.

Так как оси Ox можно придать любое направление, то такие же преобразования сдвига возможны вдоль любого диаметра. Комбинируя их с вращениями вокруг центра круга и отражениями в диаметре, получим все «движения», как они понимаются на модели; они составляются из сдвигов, вращений и отражений. Подробнее эти преобразования будут рассмотрены в следующем параграфе, где будет строго доказано, что действительно в нашей модели выполняется геометрия Лобачевского и что, в частности, преобразования, принятые за движения, удовлетворяют всем условиям (аксиомам), каким подчиняются движения в геометрии.

Повторим снова, какую же модель геометрии Лобачевского предложил Клейн. За плоскость принимается внутренность круга; точкой считается точка, прямой — хорда (концы исключены), движением считается преобразование, переводящее круг сам в себя и хорды в хорды; расположение точек (точка лежит на прямой; точка лежит между двумя другими) понимается в обычном смысле. Закон измерения длин и углов (а также и площадей) уже вытекает из того, что движение определено, и тем самым определено равенство отрезков и углов (и любых фигур), и тем же определена операция откладывания одного отрезка вдоль другого. *При всех этих условиях всякой теореме геометрии Лобачевского на плоскости соответствует факт эвклидовой геометрии внутри круга, и обратно: всякий такой факт пересказывается в виде теоремы геометрии Лобачевского.*

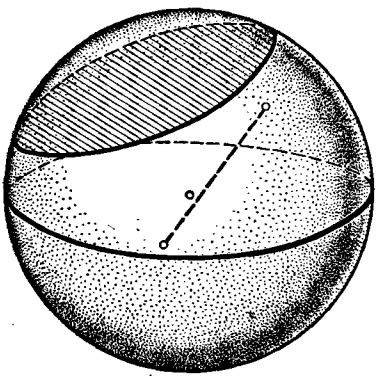


Рис. 13.

Совершенно аналогично строится модель геометрии Лобачевского в пространстве. За пространство принимается внутренность какого-либо шара (рис. 13), прямой считается хорда, плоскостью — круг с окружностью на поверхности шара, причем самая поверхность шара, а значит, концы хорд и окружности названных кругов исключаются; наконец, движение определяется как преобразование шара самого в себя, переводящее хорды в хорды.

Когда эта модель геометрии Лобачевского была дана, было установлено тем самым, что эта геометрия имеет простой реальный смысл. Геометрия Лобачевского истинна уже потому, что может быть понимаема как особое изложение геометрии в круге или в шаре. Этим же была доказана ее непротиворечивость: выводы ее не могут приводить к противоречию, так как всякий ее вывод можно пересказать на язык

обычной эвклидовой геометрии внутри круга (или внутри шара, если речь идет о геометрии Лобачевского в пространстве)[1].

3. После Клейна другую модель геометрии Лобачевского дал французский математик Пуанкаре, который применил ее к выводу важных результатов в теории функций комплексной переменной[2]. Таким образом, в его руках геометрия Лобачевского привела к решению трудных проблем из совсем другой области математики. Геометрия Лобачевского нашла ряд других приложений в математике и теоретической физике; так, например, в 1913 г., физик Варичак дал ее применения в теории относительности.

Геометрия Лобачевского успешно развивается, в ней разрабатывается теория геометрических построений, общая теория кривых и поверхностей, теория выпуклых тел и др.

§ 5. АКСИОМЫ ГЕОМЕТРИИ, ИХ ПРОВЕРКА ДЛЯ УКАЗАННОЙ МОДЕЛИ

1. Для того чтобы строго математически доказать, что модель Клейна действительно дает истолкование геометрии Лобачевского, нужно прежде всего точно формулировать, что же собственно нужно доказывать. Проверять подряд теоремы Лобачевского было бы бессмысленно; их много, притом неограниченно много, поскольку можно доказывать все новые и новые теоремы. Достаточно будет, однако, показать, что в модели Клейна выполняются основные положения геометрии Лобачевского, из которых остальные уже могут быть выведены. Но в таком случае нужно точно формулировать эти основные положения.

Таким образом, задача о доказательстве непротиворечивости геометрии Лобачевского приводит к задаче о точной и полной формулировке ее основных положений, т. е. аксиом. А так как посылки геометрии Лобачевского отличаются от посылок геометрии Эвклида одной аксиомой параллельности, то задача сводится к точной и полной формулировке аксиом эвклидовой геометрии. У Эвклида такой формулировки еще не было; у него, в частности, вовсе отсутствовало какое бы то ни было определение свойств движения или наложения фигур, хотя он ими, конечно, пользовался. Задача об уточнении и пополнении

---

[1] Математики обычно говорят, что геометрия Лобачевского изображается в геометрии Эвклида, и тем самым она непротиворечива в той же мере, в какой непротиворечива геометрия Эвклида.

[2] Модель Пуанкаре сводится к тому, что за плоскость Лобачевского принимается опять внутренность круга, но прямыми считаются уже дуги окружностей, перпендикулярные окружности данного круга; движением считается любое конформное преобразование круга самого в себя. (Связь с конформными преобразованиями и дает связь с теорией функций комплексного переменного.)

аксиом Эвклида встала во весь рост именно в связи с развитием геометрии Лобачевского, а также в связи с наметившимся в конце прошлого столетия общим течением к уточнению основ математики.

В результате исследований ряда геометров вопрос о формулировке аксиом геометрии был решен.

Вообще аксиомы можно выбирать различно, принимая в качестве основных разные понятия. Мы приведем здесь список аксиом геометрии на плоскости, в котором основными понятиями служат точка, прямая, движение и такие понятия, как точка X лежит *на* прямой a; точка B лежит *между* точками A и C; *движение* переводит точку X в точку Y. (В таком случае другие понятия через них определяются; так, например, отрезок определяется как множество всех точек, лежащих между двумя данными.)

Аксиомы делят на пять групп.

I. Аксиомы сочетания

1) Через каждые две точки проходит прямая и притом только одна.

2) На каждой прямой есть по крайней мере две точки.

3) Существуют по крайней мере три точки, не лежащие на одной прямой.

II. Аксиомы порядка

1) Из любых трех точек прямой только одна лежит между двумя другими.

2) Если A, B — две точки прямой, то на той же прямой есть хотя бы одна точка C, такая, что B лежит между A и C.

3) Прямая делит плоскость на две полуплоскости (т. е. разбивает все не лежащие на ней точки плоскости на два класса так, что точки одного класса соединимы отрезком, ее не пересекающим, а разных — нет).

III. Аксиомы движения

(Движение понимается не как преобразование отдельной фигуры, а как преобразование всей плоскости.)

1) Движение переводит прямые в прямые.

2) Два движения, произведенные одно за другим, равносильны некоторому одному движению.

3) Пусть A, A' и a, a' — две точки и исходящие из них полупрямые, а α, α' — полуплоскости, ограниченные продолженными прямыми a и a'; существует и притом единственное движение, переводящее A в A', a в a' и α в α'. (Говоря наглядно, точка A переводится в A'

переносом, затем поворотом полупрямая a переводится в a', и тогда полуплоскость α либо совпадает с α', либо еще нужно будет произвести «переворот» вокруг прямой a.)

IV. Аксиома непрерывности

Пусть точки X_1, X_2, X_3, ... расположены на прямой так, что каждая следующая лежит правее предыдущей, но при этом есть точка A, которая лежит правее их всех[1]. Тогда существует такая точка B, которая тоже лежит правее всех точек X_1, X_2, ..., но так, что сколь угодно близко от нее есть точки X_n (т. е. какую точку C левее B ни взять, на отрезке CB есть точки X_n).

V. Аксиома параллельности (Эвклида)

Через данную точку может проходить лишь одна прямая, не пересекающая данной прямой.

Таковы аксиомы, достаточные для построения эвклидовой геометрии на плоскости. Из них фактически можно вывести все теоремы школьного курса планиметрии, хотя вывод этот очень кропотлив.

Аксиомы геометрии Лобачевского отличаются только в части аксиомы параллельности.

V′. Аксиома параллельности (Лобачевского)

Через лежащую вне прямой точку проходят по крайней мере две **прямые, не пересекающие данной прямой.**

Может показаться немного странным, что в списке аксиом есть, например, такая: «на каждой прямой есть по крайней мере две точки». Ведь по нашему представлению о прямой на ней есть даже бесконечное множество точек. Не мудрено, что ни Эвклиду, ни кому-либо из математиков до конца прошлого века не приходило в голову формулировать такую аксиому: она подразумевалась. Но теперь положение изменилось. Когда мы даем новое истолкование геометрии, то под прямой понимается уже не обычная прямая, а что-то другое: геодезическая линия на поверхности, хорда круга или еще что-нибудь. Поэтому встает задача точно и исчерпывающим образом формулировать явно *все*, что мы должны потребовать от тех объектов, которые будут изображать прямые. То же относится ко всем другим понятиям и аксиомам.

Таким образом, как это уже было сказано, появление разных толкований геометрии служит одним из важных стимулов к уточнению ее основных положений. Исторически так и было: точная формулировка аксиом появилась позже моделей Бельтрами, Клейна и Пуанкаре.

---

[1] «Правее» можно соответственно заменить на «левее».

2. Теперь мы докажем, что в модели Клейна выполняются все перечисленные аксиомы, кроме эвклидовой аксиомы параллельности. Как уже было отмечено в предыдущем параграфе (рис. 11), здесь явно выполняется не она, а аксиома Лобачевского. Остается проверить аксиомы I—IV.

Плоскостью в модели служит внутренность круга (радиус его будем считать равным единице). Роль точек играют точки, роль прямых — хорды; понятия «точка лежит на прямой» и «точка лежит между двумя другими» понимаются в обычном смысле. Отсюда очевидно, что аксиомы сочетания, порядка и непрерывности выполняются. Так, например, третья аксиома порядка просто означает, что хорда делит круг на две части.

Остается проверить аксиомы движения. Движение определяется как преобразование, которое переводит круг сам в себя и прямые в прямые. Из этого определения очевидно, что эти преобразования удовлетворяют двум первым аксиомам движения: первой аксиоме — потому, что прямыми как раз считаются хорды и, стало быть, сохранение хорд означает сохранение прямых; второй аксиоме — потому, что если произвести два преобразования, переводящие круг в круг и хорды в хорды, то результирующее преобразование тем самым переведет круг в круг и хорды в хорды, т. е. будет одним из принятых за «движения».

Таким образом, остается лишь третья аксиома движения, и ее проверка представляет единственную имеющуюся здесь трудность.

Прежде всего заметим, что эта аксиома содержит два утверждения.

Пусть A, A' — две точки, a, a' — две исходящие из этих точек полупрямые, α, α' — две полуплоскости, ограниченные прямыми a, a'.

Первое утверждение состоит в том, что *существует* движение переводящее A в A', a в a', α в α'.

Второе утверждение состоит в том, что такое движение *только одно*.

Можно было бы сослаться на то, что оба эти утверждения уже доказаны в главе III, § 14 (см. том 1), но мы предпочитаем привести здесь их доказательство, не связывая его, как в главе III, с другими, более общими вопросами

Докажем для модели (т. е. в принятом понимании терминов «полупрямая», «полуплоскость», «движение») справедливость первого утверждения.

Допустим сначала, что точка A' лежит в центре круга. Выберем оси координат так, чтобы начало их лежало в центре круга, а ось Ox шла через точку A (рис. 14).

В предыдущем параграфе мы рассматривали преобразование

$$x' = \frac{x + a}{1 + ax}, \quad y' = \frac{y\sqrt{1 - a^2}}{1 + ax}. \tag{6}$$

Там было доказано, что оно есть «движение» (т. е. переводит данный круг сам в себя и прямые — в прямые).

Пусть x_0 — абсцисса точки A, ордината ее $y_0 = 0$. Поэтому, если мы возьмем $a = -x_0$, то, согласно формулам (6), точка A перейдет в точку с координатами (0, 0), т. е. в точку A'.

Так как прямые переходят при этом в прямые, то «полупрямая» (т. е. отрезок хорды) a примет некоторое положение a'' (рис. 14). Поворотом вокруг центра можно будет перевести теперь a'' в a'. «Полуплоскостью» α является один из сегментов, ограниченных «прямой» (хордой) a. Если он после движения совпадает с α', то преобразование кончено; если же не совпадает, то переворотом (отражением в диаметре a') переведем его в полукруг α'.

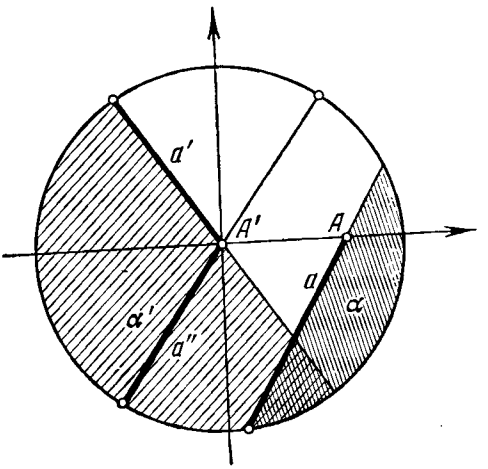


Рис. 14.

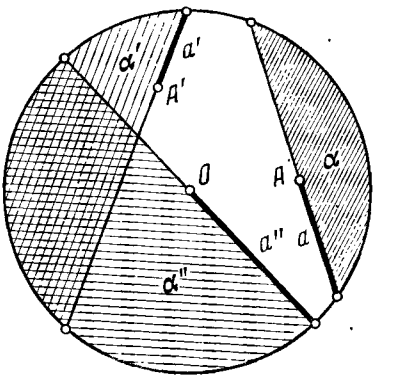


Рис. 15.

Таким образом, комбинируя «смещение» (6) с вращением и, если нужно, с отражением, мы перевели A, a, α в A', a', α'. Но результирующее всех этих «движений» тоже будет «движением»; это «движение» и переводит A, a, α в A', a', α', т. е. существование требуемого движения доказано.

Пока мы ограничивались частным случаем, когда точка A' лежит в центре. Допустим теперь, что она занимает любое положение Тогда, согласно только что доказанному, мы сможем перевести ее в центр некоторым «движением», которое обозначим D_1. При этом «полупрямая» a' перейдет в какую-то «полупрямую» a'', идущую из центра, а «полуплоскость» α' — в некоторую «полуплоскость» (полукруг) α'' (рис. 15).

Как уже доказано, мы можем посредством некоторого «движения» D_2 перевести и точку A в центр, «полупрямую» a — в a'', «полуплоскость» α — в α''. Наконец, «движением», обратным D_1, мы переведем A' на прежнее место и вместе с тем a'', α'' вернутся в исходные положения a', α'.[1]

[1] «Движение», обратное D_1, изображается формулами (4) (§ 4), если D_1 изображается формулами (3).

Таким образом, в результате комбинации «движения» D_2 и «движения», обратного D_1, мы переводим A, a, α в A', a', α'. Но комбинация «движений» есть снова «движение»; тем самым установлено, что существует движение, переводящее A, a, α в A', a', α' уже при любом положении точек A и A' в круге. Этим первое утверждение, заключенное в третьей аксиоме движения, доказано в полном объеме.

Докажем теперь, что второе утверждение также выполняется в нашей модели. Согласно смыслу этого утверждения нужно доказывать следующее.

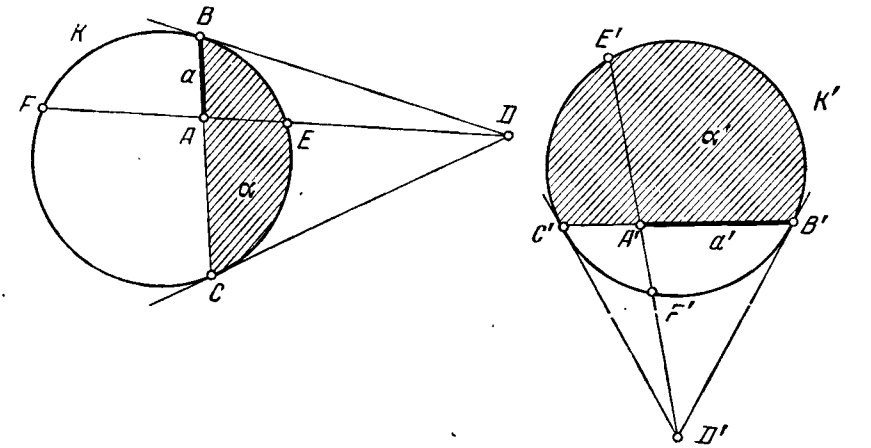


Рис. 16.

Пусть A, A' — две точки внутри круга; a, a' — исходящие из них отрезки хорд; α, α' — части круга, ограниченные этими хордами. Для ясности мы изображаем эти точки, хорды и части круга на двух разных чертежах (рис. 16), хотя, конечно, они лежат в одном рассматриваемом круге. Утверждается, что «движение», переводящее соответственно A в A', a в a', α в α', — единственное, т. е. оно вполне определяется этими данными.

При доказательстве мы будем рассматривать преобразование не только данного круга, но всей плоскости[1]. «Движение», согласно определению, переводит прямые в прямые. Преобразование, обладающее таким свойством, называется *проективным*. Стало быть, можно сказать, что «движением» у нас считается проективное преобразование, переводящее данный круг сам в себя. (На рис. 16 мы изображаем это так, что круг K переходит в круг K'. Нужно только параллельным переносом мысленно наложить круг K' на K.)

[1] Можно доказать, что преобразование круга в себя, переводящее прямые в прямые, однозначно распространяется с сохранением этого свойства на всю проективную плоскость, т. е. плоскость с присоединением бесконечно удаленных точек.

Проективные преобразования были рассмотрены в главе III (том 1), § 12, и здесь мы воспользуемся следующей доказанной там важной теоремой: проективное преобразование вполне определяется тем, куда переходят четыре точки, не лежащие по три на одной прямой.

Обратимся к рассматриваемому «движению». Оно переводит отрезок хорды a в a', а поэтому переводит точку B в B'. Так как оно переводит хорды в хорды, то оно переводит также точку C в C'.

Далее, так как «движение» вообще переводит прямые в прямые и данный круг сам в себя, а в нашем изображении круг K — в круг K', то касательные в точках B и C переходят в касательные в точках B' и C'. Поэтому точка D пересечения первых касательных переходит в точку D' пересечения вторых касательных[1] (рис. 16).

Так как точка A к тому же переходит в A' и прямые переходят в прямые, то прямая AD переходит в прямую $A'D'$. Прямая AD пересекает окружность нашего круга в точках E, F, а прямая $A'D'$ пересекает ее в точках E', F'. (На чертеже эти точки лежат на окружностях K и K'.) Так как круг переходит сам в себя, то точки E, F переходят в точки E', F'. Пусть точка E лежит на дуге, ограничивающей часть α, а E' — на дуге, ограничивающей часть α'. Тогда, поскольку α по условию переходит в α', то E переходит именно в E' и соответственно F в F'.

Итак, мы получили, что при рассматриваемом «движении» точки B, C, E, F на окружности переходят в точки B', C', E', F'. Точки B, C, E, F, так же как, очевидно, и точки B', C', E', F', не лежат по три на одной прямой. Поэтому, согласно приведенной выше теореме, проективное преобразование, переводящее B, C, E, F в B', C', E', F', — единственное. Но «движение» и есть проективное преобразование. Стало быть, наше «движение», переводящее A, a, α в A', a', α', единственное, что и требовалось доказать.

Таким образом, мы доказали, что в рассматриваемой модели действительно выполняются все аксиомы эвклидовой геометрии, кроме аксиомы параллельности, или, иными словами, что в модели выполняются все аксиомы геометрии Лобачевского. Модель, следовательно, на самом деле реализует геометрию Лобачевского. Эта геометрия как бы сведена к геометрии Эвклида внутри круга, изложенной особым образом, с тем условным пониманием терминов «прямая» и «движение», которое принято в модели. Кстати, это позволяет уже развивать геометрию Лобачевского на данной конкретной модели, что во многих вопросах оказывается более удобным.

С точки зрения логического анализа оснований геометрии приведенное доказательство показывает, во-первых, что геометрия Лобачев-

[1] Если, например, касательные в точках B и C параллельны, то точка D' есть «бесконечно удаленная».

ского непротиворечива, а во-вторых, что постулат о параллельных заведомо не может быть выведен из перечисленных выше остальных аксиом.

§ 6. ВЫДЕЛЕНИЕ САМОСТОЯТЕЛЬНЫХ ГЕОМЕТРИЧЕСКИХ ТЕОРИЙ ИЗ ЭВКЛИДОВОЙ ГЕОМЕТРИИ

1. Принципиальное развитие геометрии параллельно с созданием геометрии Лобачевского шло еще по другому пути. В богатстве всех геометрических свойств пространства выделялись и подвергались самостоятельному изучению отдельные группы свойств, отличающиеся своеобразной замкнутостью и устойчивостью. Такие, обособленные по своим методам, исследования составили новые главы геометрии — науки о пространственных формах, подобно тому, как, например, анатомия или физиология составляют различные главы науки об организме человека.

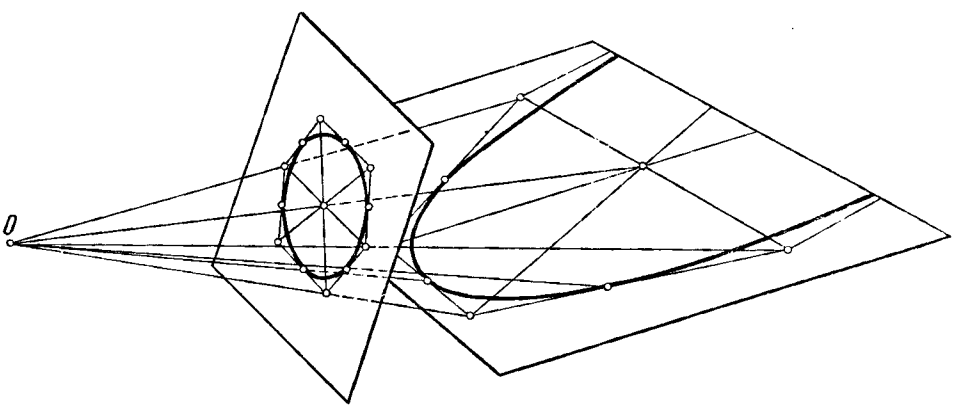


Рис. 17.

Первоначально геометрия вообще не расчленялась. Она изучала главным образом метрические — связанные с измерением размеров фигур — свойства пространства. Лишь попутно рассматривались обстоятельства, связанные не с измерением, а с качественным характером взаимного расположения фигур, причем уже давно замечали, что часть таких свойств отличается своеобразной устойчивостью, сохраняясь при довольно существенных искажениях формы и изменениях положения фигур.

Рассмотрим, например, проектирование рисунка с одной плоскости на другую (рис. 17). Длины отрезков при этом меняются, изменяются углы, явно искажаются очертания предметов. Однако, например, свойство ряда точек лежать на одной прямой сохраняется, сохраняется свойство прямой касаться какой-нибудь линии и т. д.

О проектировании и проективных преобразованиях уже шла речь в главе III, где отмечалась их очевидная связь с перспективой — изображением пространственных фигур на плоскости. Исследование

свойств перспективы восходит к далеким временам до Эвклида, к работам древних архитекторов; перспективой занимались художники: Дюрер, Леонардо да Винчи, инженер и математик Дезарг (XVII в.). Наконец, в начале XIX в. Понселе впервые последовательно выделил и исследовал геометрические свойства, сохраняющиеся при любых проективных преобразованиях плоскости (или пространства), создав тем самым самостоятельную науку — *проективную геометрию*[1].

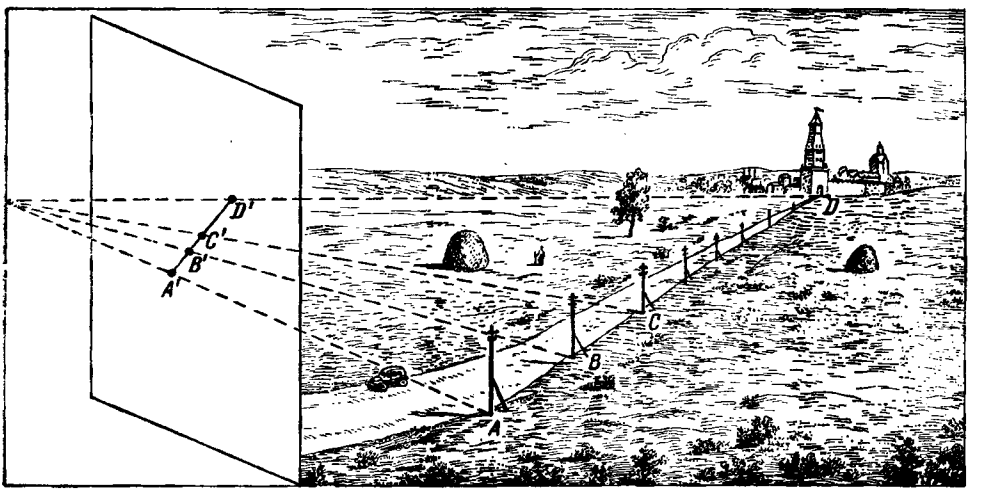


Рис. 18.

Казалось бы, свойств, сохраняющихся при любом проективном преобразовании, немного и они весьма примитивны, но это далеко не так. Не сразу, например, обращаешь внимание на то, что теорема, утверждающая, что точки попарного пересечения продолжений противоположных сторон вписанного в круг шестиугольника лежат на одной прямой, верна и для эллипса, параболы и гиперболы. Она говорит лишь о проективных свойствах, а эти кривые получаются проектированием круга. Тем более не очевидно, что теорема о пересечении в одной точке диагоналей описанного шестиугольника является своеобразным аналогом указанной выше теоремы; их глубокая связь вскрывается именно в проективной геометрии. Не очевидно также, что при проектировании, несмотря на искажение расстояний, для всяких четырех точек A, B, C, D (рис. 18), лежащих на одной прямой, двойное отношение $\frac{AC}{CB} : \frac{AD}{DB}$ остается неизменным

$$\frac{AC}{CB} : \frac{AD}{DB} = \frac{A'C'}{C'B'} : \frac{A'D'}{D'B'}.$$

[1] Понселе — французский военный инженер, занимавшийся геометрическими исследованиями во время пребывания в плену в России после 1812 г. Свой «Трактат о проективных свойствах фигур» он опубликовал в 1822 г.

Это влечет за собой соблюдение в перспективе многих соотношений. Например, используя это обстоятельство, легко по снимку уходящей в даль дороги (рис. 18), где видно расположение телеграфных столбов A', B', C', определить расстояние от них до пункта D'.

О проективной геометрии и использовании ее выводов в работах по аэрофотосъемке шла речь в главе III (том 1). Разумеется, ее законы используются в архитектуре, а также при построении панорам, декораций и т. п.

Выделение проективной геометрии сыграло важную роль в развитии самой геометрии.

2. Другим примером самостоятельной геометрии может служить *аффинная геометрия*. В ней исследуются те свойства фигур, которые не меняются при любых преобразованиях, в которых декартовы координаты первоначального (x, y, z) и нового (x', y', z') положения каждой точки связаны между собой линейными уравнениями:

$$\begin{aligned} x' &= a_1x + b_1y + c_1z + d_1, \\ y' &= a_2x + b_2y + c_2z + d_2, \\ z' &= a_3x + b_3y + c_3z + d_3 \end{aligned}$$

(предполагается, что определитель

$$\begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix}$$

отличен от нуля).

Как оказывается, любое аффинное преобразование сводится к движению, может быть, еще отражению в плоскости, а затем к сжатиям или растяжениям пространства в трех взаимно перпендикулярных направлениях.

При любом из таких преобразований сохраняются уже довольно многие свойства фигур. Прямые остаются прямыми (вообще все «проективные» свойства сохраняются); кроме того, параллельные прямые остаются параллельными; сохраняется отношение объемов, отношение площадей фигур, лежащих в параллельных плоскостях или на одной и той же плоскости, отношение длин отрезков, лежащих на одной прямой или параллельных прямых, и т. д. Многие хорошо известные всем теоремы по существу принадлежат аффинной геометрии. Таковы, например, утверждения, что медианы треугольника пересекаются в одной точке, что диагонали параллелограма взаимно делятся пополам, что середины параллельных хорд эллипса лежат на одной прямой и т. д.

С аффинной геометрией теснейшим образом связана вся теория кривых (и поверхностей) 2-го порядка. Само разделение этих кривых на эллипсы, параболы, гиперболы основано между прочим на аффинных свойствах фигуры: при аффинных преобразованиях эллипс преоб-

разуется именно в эллипс, но никак не в параболу или гиперболу; аналогично, парабола может быть преобразована в любую другую параболу, но не в эллипс и т. д.

Важность выделения и подробного исследования общих аффинных свойств фигур усугубляется тем, что несравненно более сложные преобразования в бесконечно малом оказываются по существу линейными, т. е. аффинными, а применение методов дифференциального исчисления связано именно с рассмотрением бесконечно малых областей пространства.

3. В 1872 г. Клейн в лекции, прочитанной в университете в Эрлангене и известной поэтому под названием Эрлангенской программы, суммируя результаты развития проективной, аффинной и других «геометрий», четко сформулировал общий принцип их построения, именно: можно рассматривать любую группу взаимно однозначных преобразований пространства и исследовать те свойства фигур, которые сохраняются при преобразованиях этой группы[1].

С этой точки зрения свойства пространства как бы расслаиваются по их глубине и устойчивости. Обычная эвклидова геометрия была создана путем отвлечения от всех свойств реальных тел, кроме геометрических; здесь, в специальных разделах геометрии, мы как бы совершаем еще одно абстрагирование уже внутри геометрии, отвлекаясь от ряда геометрических свойств, кроме определенного их круга, интересующего нас в данной отрасли.

По указанному Клейном принципу можно строить много геометрий. Например, можно рассматривать преобразования, сохраняющие углы между любыми линиями (конформные преобразования пространства), и, исследуя сохраняющиеся при таких преобразованиях свойства фигур, говорить о соответствующей конформной геометрии. Можно рассматривать преобразования не обязательно всего пространства. Так, рассматривая точки и хорды круга при всех его преобразованиях в себя, переводящих хорды в хорды, и выделяя свойства, сохраняющиеся при таких преобразованиях, мы получаем геометрию, которая, как было показано в §§ 4 и 5, совпадает с геометрией Лобачевского.

4. Развитие выделившихся таким образом теорий даже с принципиальной стороны (не говоря уже о фактическом содержании) не остановилось на том, что здесь было сказано.

Интересуясь, например, только аффинными свойствами фигур, мы можем, отвлекаясь от всех других свойств, мыслить себе пространство,

---

[1] Слово «группа» используется здесь не просто в собирательном смысле. Говоря о группе преобразований (см. главу XX), имеют в виду такую совокупность преобразований, в которой заведомо содержится тождественное преобразование (оставляющее все точки на месте), где наряду с каждым нетождественным преобразованием имеется и ему обратное (возвращающее все точки на прежнее место) и где вместе с каждыми двумя преобразованиями этой совокупности содержится и преобразование, равносильное этим двум, последовательно произведенным.

а также геометрические фигуры в нем, обладающими *только* интересующими нас свойствами и как бы вообще не имеющими никаких других свойств. В этом «пространстве» фигуры вообще не будут иметь никаких других свойств, кроме аффинных. Естественно будет пытаться геометрию такого абстрактного пространства тоже излагать аксиоматически, т. е. считать, что речь идет о некоторых отвлеченных объектах: «точках», «прямых», «плоскостях», свойства которых (этих свойств будет, очевидно, меньше, чем в случае эвклидовой геометрии) указываются в некоторых аксиомах, причем выводимые из этих аксиом следствия будут соответствовать аффинным свойствам фигур обычного пространства.

Это действительно удается сделать; и такую совокупность абстрактных «точек», «прямых» и «плоскостей» с системой их свойств называют *аффинным пространством.*

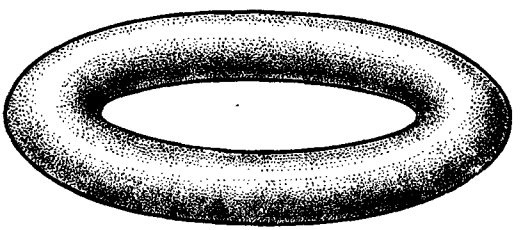


Рис. 19.

Точно так же можно мыслить себе абстрактную систему объектов, обладающих только тем кругом свойств, которые соответствуют проективным свойствам фигур эвклидова пространства. (На этот раз отличие системы аксиом от аксиом обычной геометрии оказывается еще более существенным.)

5. Если дальше углубляться в природу геометрических образов, можно заметить, что в целом ряде вопросов играют роль свойства, еще более глубокие, чем проективные, и настолько прочно связанные с данной фигурой, что они сохраняются при любых искажениях фигур, если только эти искажения не приводят к разрыву или склеиванию частей фигуры. (Желая уточнить представление о таком непрерывном искажении, можно, кроме наглядного описания, сослаться на известное нам из анализа определение непрерывной функции и сказать, что речь идет о любом преобразовании всех точек фигуры в новое положение, при котором декартовы координаты точек в новом положении выражаются непрерывными функциями их прежних координат, а старые координаты в свою очередь могут быть выражены как непрерывные функции новых.)

Свойства фигур, сохраняющиеся при любых таких преобразованиях, называются топологическими, а наука, изучающая их, — *топологией* (см. главу XVIII).

Тесно связанные с фигурами топологические свойства для простейших фигур отличаются исключительной наглядностью. Почти очевидно, например, что на плоскости всякая линия, которую можно получить, непрерывно деформируя окружность, разбивает плоскость на две части, лежащие внутри и вне этого, сколь угодно извилистого контура; поэтому

свойство окружности делить плоскость — топологическое. Наглядно, пожалуй, очевидно, что поверхность тора (рис. 19) никак нельзя превратить непрерывным преобразованием в сферу, поэтому свойство какой-либо поверхности допускать непрерывное преобразование, скажем в поверхность тора, будет ее топологическим свойством, отличающим ее от многих других поверхностей.

Рассуждения, связанные с непрерывностью, отличаются наглядностью и часто столь хорошо поясняют суть дела, что представляется весьма заманчивым уметь превращать их в строгие доказательства, а тем более распространять аналогичные приемы на другие несравненно более сложные задачи.

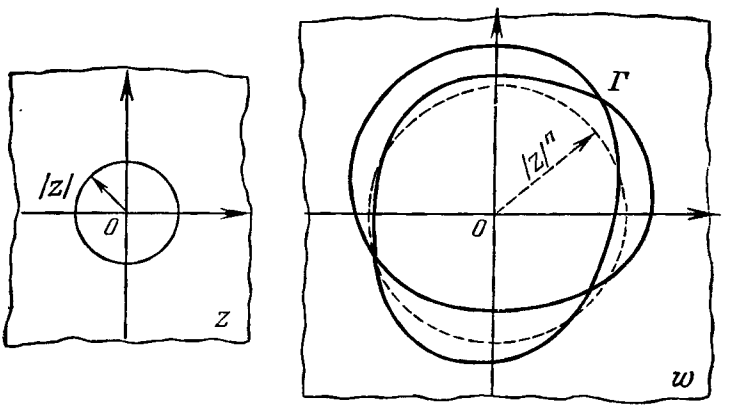


Рис. 20.

Вот, например, рассуждение, поясняющее справедливость основной теоремы алгебры о том, что всякое уравнение

$$z^n + a_1 z^{n-1} + a_2 z^{n-2} + \ldots + a_{n-1} z + a_n = 0 \qquad (7)$$

имеет хотя бы один действительный или комплексный корень.

Пусть z — точки одной комплексной плоскости, а $w = f(z)$ — соответствующие им точки другой комплексной плоскости w, причем $f(z)$ обозначает левую часть уравнения (7). При z очень больших по абсолютной величине функция $f(z)$ относительно мало отличается от z^n; функция же z^n весьма проста. В частности, легко проверить, что если точка z, непрерывно двигаясь, опишет на комплексной плоскости окружность с центром в начале координат, то точка $w_1 = z^n$ ровно n раз обойдет аналогичную окружность радиуса $|z|^n$ на плоскости w [1]. Точка же

[1] В самом деле, если $z = \rho(\cos\varphi + i\sin\varphi)$, то, как известно [см. главу IV (том 1), § 3], $z^n = \rho^n(\cos\varphi n + i\sin\varphi n)$; поэтому, с изменением аргумента φ числа z от 0 до 2π, аргумент числа z^n будет изменяться от 0 до $2\pi n$, т. е. идущий в точку z^n радиус сделает при ее движении n полных оборотов.

$w=f(z)$ опишет n петель, образующих какой-то контур Г, сравнительно близкий к линии, пройденной точкой $w_1=z^n$ (рис. 20).

Если теперь окружность, описываемую точкой z, непрерывно стягивать в одну точку, то и n раз запетленный контур Г, описываемый точкой $w=f(z)$, будет непрерывно деформироваться, также стягиваясь в точку. Но довольно очевидно, что он не может стянуться в точку, ни разу не пересекая при этом начало координат O, которое этот контур первоначально охватывал. Значит, он хотя бы раз пройдет через O, и при таком z будет $w=f(z)=0$. Это z и будет корнем уравнения (7). Собственно говоря, уже ясно, что корней в известном смысле должно быть именно n, так как каждая из n петель контура Г, стягиваясь, пересечет точку O.

Наше рассуждение требует, конечно, уточненного обоснования тех, как раз топологических, свойств контура и его деформации, которыми мы здесь воспользовались.

Можно было бы привести много примеров использования топологических свойств в самых различных, порой весьма далеких от геометрии разделах математики.

При исследовании топологических свойств перед нами опять встает возможность мыслить себе отвлеченную совокупность объектов, обладающих вообще только такого рода свойствами (см. § 7 главы XX). Такую совокупность называют *абстрактным топологическим пространством*.

Эта точка зрения уже несравненно шире, чем исследование топологических свойств именно геометрических фигур. Топологические пространства могут быть весьма разные; скажем, все точки поверхности тора с их общей закономерностью взаимного прилегания образуют одно топологическое пространство, все точки плоскости — другое, все эвклидово пространство — третье; все точки многолистных римановых поверхностей, о которых шла речь в главе IX (том 2), § 5, посвященной теории функций комплексного переменного, образуют другие, притом разные, топологические пространства. Но замечательнее всего, что между объектами, далеко не похожими на наше представление о геометрических точках, часто ясно устанавливается понятие близости и прилегания. Скажем, для всевозможных положений какого-нибудь шарнирного механизма можно ясно указать, что значит «близкие» положения, что значит одно положение «прилегает» к бесконечной серии других, среди которых есть положения, сколь угодно близкие к данному.

Мы видим, что понятие топологического пространства является чрезвычайно общим. В этой связи мы еще раз вернемся к нему в § 8.

Целью этого параграфа было не столько дать читателю представление о разных геометриях, сколько стремление показать, что самые конкретные задачи приводят к выделению и исследованию отдельных групп геометрических свойств; что их исследование влечет за собой создание представления об отвлеченных геометрических объектах,

обладающих *только* этими свойствами, т. е. что выделение этих свойств в их чистом виде приводит нас к представлению о соответствующем абстрактном пространстве.

О других путях, также приводящих к построению разного рода абстрактных пространств, будет идти речь в следующих параграфах.

§ 7. МНОГОМЕРНОЕ ПРОСТРАНСТВО

1. Важным этапом в развитии новых геометрических идей было создание геометрии многомерного пространства, о котором уже шла речь в предыдущей главе. Одной из причин ее возникновения служило стремление использовать геометрические соображения при решении вопросов алгебры и анализа. Геометрический подход к решению аналитических вопросов основан на методе координат. Приведем простой пример.

Требуется узнать, сколько целочисленных решений имеет неравенство $x^2+y^2<N$. Рассматривая x и y как декартовы координаты на плоскости, видим, что вопрос сводится к следующему: сколько точек с целочисленными координатами содержится внутри круга радиуса $\sqrt{N}$. Точки с целочисленными координатами — это вершины квадратов со стороной единичной длины, покрывающих плоскость (рис. 21). Число таких точек внутри круга приближенно равно числу квадратов, лежащих внутри круга, т. е. приблизительно равно площади круга радиуса $\sqrt{N}$. Таким образом, интересующее нас число решений неравенства приближенно равно πN. При этом нетрудно доказать, что допускаемая здесь относительная ошибка стремится к нулю при $N\to\infty$. Более точное исследование этой погрешности представляет собой весьма трудную задачу теории чисел, служившую в сравнительно недавнее время предметом глубоких исследований.

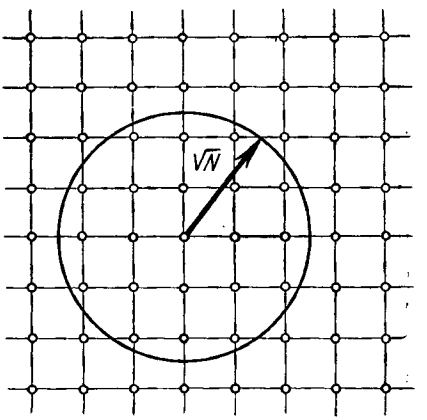


Рис. 21.

В разобранном примере оказалось достаточным перевести задачу на геометрический язык, чтобы сразу получить результат, далеко не очевидный с точки зрения «чистой алгебры». Совершенно так же решается аналогичная задача для неравенства с тремя неизвестными. Однако, если неизвестных более трех, этот метод не удается применить, поскольку наше пространство трехмерно, т. е. положение точки в нем определяется тремя координатами. Для сохранения полезной геометрической аналогии в подобных случаях вводят представление об абстрактном

«n-мерном пространстве», точки которого определяются n координатами $x_1, x_2, \ldots, x_n$. При этом основные понятия геометрии обобщаются таким образом, что геометрические соображения оказываются применимыми к решению задач с n переменными; это сильно облегчает нахождение результатов. Возможность такого обобщения основана на единстве алгебраических закономерностей, в силу которого многие задачи решаются совершенно единообразно при любом числе переменных. Это позволяет применять геометрические соображения, действующие при трех переменных, к любому их числу.

2. Зачатки понятия о четырехмерном пространстве встречаются еще у Лагранжа, который в своих работах по механике рассматривал время формально как «четвертую координату» наряду с тремя пространственными. Но первое систематическое изложение начал многомерной геометрии было дано в 1844 г. немецким математиком Грассманом и независимо от него англичанином Кэли. Они шли при этом путем формальной аналогии с обычной аналитической геометрией. Эта аналогия в современном изложении выглядит в общих чертах следующим образом.

Точка в n-мерном пространстве определяется n координатами $x_1, x_2, \ldots, x_n$. Фигура в n-мерном пространстве — это геометрическое место, или множество точек, удовлетворяющих тем или иным условиям. Например, «n-мерный куб» определяется как геометрическое место точек, координаты которых подчинены неравенствам: $a \leqslant x_i \leqslant b$ $(i = 1, 2, \ldots, n)$. Аналогия с обычным кубом здесь совершенно прозрачна: в случае, когда $n = 3$, т. е. пространство трехмерно, наши неравенства действительно определяют куб, ребра которого параллельны осям координат и длина ребер равна $b - a$ (на рис. 22 изображен случай $a = 0$, $b = 1$).

Расстояние между двумя точками можно определить как корень квадратный из суммы квадратов разностей координат

$$d = \sqrt{(x_1' - x_1'')^2 + (x_2' - x_2'')^2 + \ldots + (x_n' - x_n'')^2}.$$

Это представляет собой прямое обобщение известной формулы для расстояния на плоскости или в трехмерном пространстве, т. е. при $n = 2$ или 3.

Теперь можно определить в n-мерном пространстве равенство фигур. Две фигуры считаются равными, если между их точками можно установить такое соответствие, при котором расстояния между парами соответственных точек равны. Преобразование, сохраняющее расстояния, можно назвать обобщенным движением[1]. Тогда по аналогии с обычной

[1] Обобщение состоит не только в переходе к n измерениям, но также в том, что к движениям присоединяется еще отражение в плоскости, так как оно тоже не изменяет расстояний между точками.

эвклидовой геометрией можно сказать, что предмет n-мерной геометрии составляют свойства фигур, сохраняющиеся при обобщенных движениях. Это определение предмета n-мерной геометрии было установлено в 70-х годах и дало точную основу для ее разработки. С тех пор n-мерная геометрия служит предметом многочисленных исследований во всех направлениях, аналогичных направлениям эвклидовой геометрии (элементарная геометрия, общая теория кривых и т. п.).

Понятие расстояния между точками позволяет перенести на n-мерное пространство также другие понятия геометрии, такие как отрезок, шар, длина, угол, объем и т. п. Например, n-мерный шар определяется как множество точек, удаленных от данной не больше, чем на данное R. Поэтому аналитически шар задается неравенством

$$(x_1 - a_1)^2 + \ldots + (x_n - a_n)^2 \leqslant R^2,$$

где $a_1, \ldots, a_n$ — координаты его центра. Поверхность шара задается уравнением

$$(x_1 - a_1)^2 + \ldots + (x_n - a_n)^2 = R^2.$$

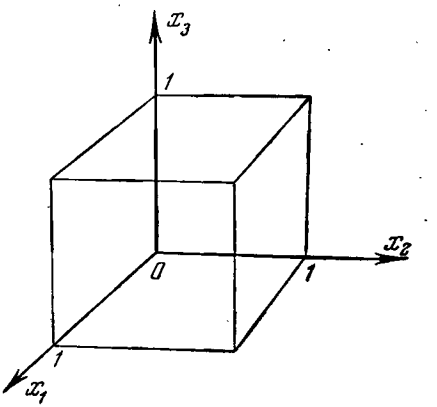


Рис. 22.

Отрезок AB можно определить как множество таких точек X, что сумма расстояний от X до A и B равна расстоянию от A до B. (Длина отрезка есть расстояние между его концами.)

3. Остановимся несколько подробнее на плоскостях различного числа измерений.

В трехмерном пространстве таковыми являются одномерные «плоскости» — прямые и обычные (двумерные) плоскости. В n-мерном пространстве при $n > 3$ вводятся в рассмотрение еще многомерные плоскости числа измерений от 3 до $n - 1$.

Как известно, в трехмерном пространстве плоскость задается одним линейным уравнением, а прямая — двумя такими уравнениями.

Путем прямого обобщения приходим к следующему определению: ***k*-мерной плоскостью** в n-мерном пространстве называется геометрическое место точек, координаты которых удовлетворяют системе $n - k$ линейных уравнений

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n + b_1 &= 0, \\ a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n + b_2 &= 0, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots & \\ a_{n-k,1}x_1 + a_{n-k,2}x_2 + \ldots + a_{n-k,n}x_n + b_n &= 0, \end{aligned} \tag{8}$$

причем уравнения совместны и независимы (т. е. ни одно из них не является следствием других). Каждое из этих уравнений представляет $(n-1)$-мерную плоскость, а все они вместе определяют общие точки $n-k$ таких плоскостей.

То, что уравнения (8) совместны, означает, что вообще есть точки, им удовлетворяющие, т. е. $n-k$ данных $(n-1)$-мерных плоскостей пересекаются. То, что ни одно уравнение не является следствием других, означает, что ни одно из них нельзя исключить. Иначе система сводилась бы к меньшему числу уравнений и определяла бы плоскость большего числа измерений. Таким образом, говоря геометрически, дело сводится к тому, что k-мерная плоскость определяется как пересечение $n-k$ штук $(n-1)$-мерных плоскостей, представляемых независимыми уравнениями. В частности, если $k=1$, то имеем $n-1$ уравнений, которые определяют «одномерную плоскость», т. е. прямую. Таким образом, данное определение k-мерной плоскости представляет естественное формальное обобщение известных результатов аналитической геометрии. Польза этого обобщения обнаруживается уже в том, что выводы, касающиеся систем линейных уравнений, получают геометрическое истолкование, которое делает эти выводы более ясными. С таким геометрическим подходом к вопросам линейной алгебры читатель мог ознакомиться в главе XVI.

Важным свойством k-мерной плоскости является то, что она может рассматриваться сама как k-мерное пространство. Так, например, трехмерная плоскость сама есть обычное трехмерное пространство. Это дает возможность переносить на пространства высшего числа измерений многие выводы, полученные для пространств низшего числа измерений, подобно обычным рассуждениям от n к $n+1$.

Если уравнения (8) совместны и независимы, то, как доказывается в алгебре, из n переменных x_i можно выбрать k так, что остальные $n-k$ переменных можно через них выразить[1]. Например:

$$\begin{aligned}
x_{k+1} &= c_{11}x_1 + c_{12}x_2 + \ldots + c_{1k}x_k + d_1,\\
x_{k+2} &= c_{21}x_1 + c_{22}x_2 + \ldots + c_{2k}x_k + d_2,\\
&\ldots\ldots\ldots\ldots\ldots\ldots\\
x_n &= c_{n-k,1}x_1 + c_{n-k,2}x_2 + \ldots + c_{n-k,k}x_k + d_k.
\end{aligned}$$

Здесь $x_1, x_2, \ldots, x_k$ могут принимать любые значения, а остальные x_i через них определяются. Это значит, что положение точки на k-мерной плоскости определяется уже k координатами, могущими принимать любые значения. Именно в этом смысле плоскость имеет k измерений.

[1] Эти k переменных из x_i, вообще говоря, нельзя выбирать произвольно. Например, из системы $x_1+x_2+x_3=0$, $x_1-x_2-x_3=0$ значение x_1 определяется однозначно: $x_1=0$, и через него x_2 и x_3, очевидно, нельзя выразить. Утверждается, однако, что необходимые k из x_i всегда можно выбрать.

Из определения плоскостей разного числа измерений можно чисто алгебраически вывести следующие основные теоремы.

1) Через каждые $k+1$ точку, не лежащую на одной $(k-1)$-мерной плоскости, проходит k-мерная плоскость и притом только одна.

Полная аналогия с известными фактами элементарной геометрии здесь очевидна. Доказательство этой теоремы опирается на теорию систем линейных уравнений и несколько сложно, так что мы не будем его излагать.

2) Если l-мерная и k-мерная плоскости в n-мерном пространстве имеют хотя бы одну общую точку и при этом $l+k \geqslant n$, то они пересекаются по плоскости размерности не меньшей, чем $l+k-n$.

Как частный случай отсюда вытекает, что две двумерные плоскости в трехмерном пространстве, если они не совпадают и не параллельны, пересекаются по прямой ($n=3$, $l=2$, $k=2$, $l+k-n=1$). Но уже в четырехмерном пространстве две двумерных плоскости могут иметь единственную общую точку. Например, плоскости, задаваемые системами уравнений:

$$\left.\begin{aligned} x_1 &= 0 \\ x_2 &= 0 \end{aligned}\right\}, \qquad \left.\begin{aligned} x_3 &= 0 \\ x_4 &= 0 \end{aligned}\right\},$$

очевидно, пересекаются в единственной точке с координатами $x_1=0$, $x_2=0$, $x_3=0$, $x_4=0$.

Доказательство сформулированной теоремы чрезвычайно просто: l-мерная плоскость задается $n-l$ уравнениями; k-мерная задается $n-k$ уравнениями; координаты точек пересечения должны удовлетворять одновременно всем $(n-l)+(n-k)=n-(l+k-n)$ уравнениям. Если ни одно уравнение не является следствием остальных, то по самому определению плоскости в пересечении имеем $(l+k-n)$-мерную плоскость; в противном случае получается плоскость большего числа измерений.

К двум указанным теоремам можно добавить еще две.

3) На каждой k-мерной плоскости есть по крайней мере $k+1$ точек, не лежащих в плоскости меньшего числа измерений. В n-мерном пространстве есть по крайней мере $n+1$ точек, не лежащих ни в какой плоскости.

4) Если прямая имеет с плоскостью (любого числа измерений) две общие точки, то она целиком лежит в этой плоскости. Вообще, если l-мерная плоскость имеет с k-мерной плоскостью $l+1$ общих точек, не лежащих в $(l-1)$-мерной плоскости, то она целиком лежит в этой k-мерной плоскости.

Заметим, что n-мерную геометрию можно строить, исходя из аксиом, обобщающих аксиомы, сформулированные в § 5. При таком подходе четыре указанные выше теоремы принимаются за аксиомы сочетания. Это кстати показывает, что понятие аксиомы относительно: одно и то же

утверждение при одном построении теории выступает как теорема, при другом — как аксиома.

4. Мы получили общее представление о математическом понятии многомерного пространства. Чтобы выяснить реальный физический смысл этого понятия, обратимся снова к задаче графического изображения. Пусть, например, мы хотим изобразить зависимость давления газа от объема. Берем на плоскости координатные оси и на одной оси откладываем объем v, а на другой — давление p. Зависимость давления от объема при данных условиях изобразится некоторой кривой (при данной температуре для идеального газа это будет гипербола согласно известному закону Бойля-Мариотта). Но если мы имеем более сложную физическую систему, состояние которой задается уже не двумя данными (как объем и давление в случае газа), а, скажем, пятью, то графическое изображение ее поведения приводит к представлению соответственно о пятимерном пространстве.

Пусть, например, речь идет о сплаве трех металлов или о смеси трех газов. Состояние смеси определяется четырьмя данными: температурой T, давлением p и процентными содержаниями c_1, c_2 двух газов (процентное содержание третьего газа определяется тогда тем, что общая сумма процентных содержаний равна 100%, так что $c_3=100-c_1-c_2$). Состояние такой смеси определяется, следовательно, четырьмя данными. Графическое его изображение требует или соединения нескольких диаграмм, или приходится представлять себе это состояние в виде точки четырехмерного пространства с четырьмя координатами T, p, c_1, c_2. Таким представлением фактически пользуются в химии; применение методов многомерной геометрии к задачам этой науки разработано американским ученым Гиббсом и школой советских физико-химиков академика Курнакова. Здесь введение многомерного пространства диктуется стремлением сохранить полезные геометрические аналогии и соображения, исходящие из простого приема графического изображения.

Приведем еще пример из области геометрии. Шар задается четырьмя данными: тремя координатами его центра и радиусом. Поэтому шар можно представлять точкой в четырехмерном пространстве. Специальная геометрия шаров, которую построили около ста лет назад некоторые математики, может рассматриваться поэтому как некоторая четырехмерная геометрия.

Из всего сказанного выясняется общее реальное основание для введения понятия многомерного пространства. Если какая-либо фигура, или состояние какой-либо системы и т. п. задается n данными, то эту фигуру, это состояние и т. п. можно мыслить как точку некоторого n-мерного пространства. Польза этого представления примерно та же, что польза обычных графиков: она состоит в возможности применить известные геометрические аналогии и методы для изучения рассматриваемых явлений.

В математическом понятии многомерного пространства нет, следовательно, никакой мистики. Оно представляет собой не более как некоторое абстрактное понятие, выработанное математиками для того, чтобы описывать на геометрическом языке такие вещи, которые не допускают простого геометрического изображения в обычном смысле. Это абстрактное понятие имеет вполне реальное основание, оно отражает действительность и было вызвано потребностями науки, а не праздной игрой воображения. Оно отражает тот факт, что существуют вещи, которые, как шар или смесь из трех газов, характеризуются несколькими данными, так что совокупность всех таких вещей является многомерной. Число измерений в данном случае есть именно число этих данных. Как точка, двигаясь в пространстве, меняет три свои координаты, так шар, двигаясь, расширяясь и сжимаясь, изменяет четыре свои «координаты», т. е. четыре величины, которые его определяют.

В следующих параграфах мы еще остановимся на многомерной геометрии. Сейчас же важно только понять, что она является методом математического описания реальных вещей и явлений. Представление о каком-то четырехмерном пространстве, в котором находится наше реальное пространство — представление, использовавшееся некоторыми беллетристами и спиритами, не имеет отношения к математическому понятию о четырехмерном пространстве. Если и можно говорить здесь об отношении к науке, то разве лишь в смысле фантастического искажения научных понятий.

5. Как уже говорилось, геометрия многомерного пространства строилась сначала путем формального обобщения обычной аналитической геометрии на произвольное число переменных. Однако такой подход к делу не мог полностью удовлетворить математиков. Ведь цель состояла не столько в обобщении геометрических понятий, сколько в обобщении самого геометрического метода исследования. Поэтому важно было дать чисто геометрическое изложение n-мерной геометрии, не зависящее от аналитического аппарата. Впервые это было сделано швейцарским математиком Шлефли в 1852 г., рассмотревшим в своей работе вопрос о правильных многогранниках многомерного пространства. Правда, работа Шлефли не была оценена современниками, так как для ее понимания нужно было в той или иной мере подняться до абстрактного взгляда на геометрию. Лишь дальнейшее развитие математики внесло в этот вопрос полную ясность, выяснив исчерпывающим образом взаимоотношение аналитического и геометрического подходов. Не имея возможности углубляться в этот вопрос, мы ограничимся примерами геометрического изложения n-мерной геометрии. Рассмотрим геометрическое определение n-мерного куба. Двигая отрезок в плоскости перпендикулярно самому себе на расстояние, равное его длине, мы зачертим квадрат, т. е. двумерный куб (рис. 23, *а*). Совершенно аналогично, двигая квадрат в направлении, перпендикулярном его плоскости, на расстояние, равное его

стороне, мы зачертим трехмерный куб (рис. 23, *б*). Чтобы получить четырехмерный куб, применяем то же построение: взяв в четырехмерном пространстве трехмерную плоскость и в ней трехмерный куб, двигаем его в направлении, перпендикулярном этой трехмерной плоскости, на расстояние, равное ребру (по определению прямая перпендикулярна k-мерной плоскости, если она перпендикулярна всякой прямой, лежащей в этой плоскости). Это построение условно представлено на рис. 23, *в*. Здесь изображено два трехмерных куба Q и Q' — данный куб в первоначальном и конечном положении. Линии, соединяющие вершины этих кубов, изображают те отрезки, по которым двигаются вершины при

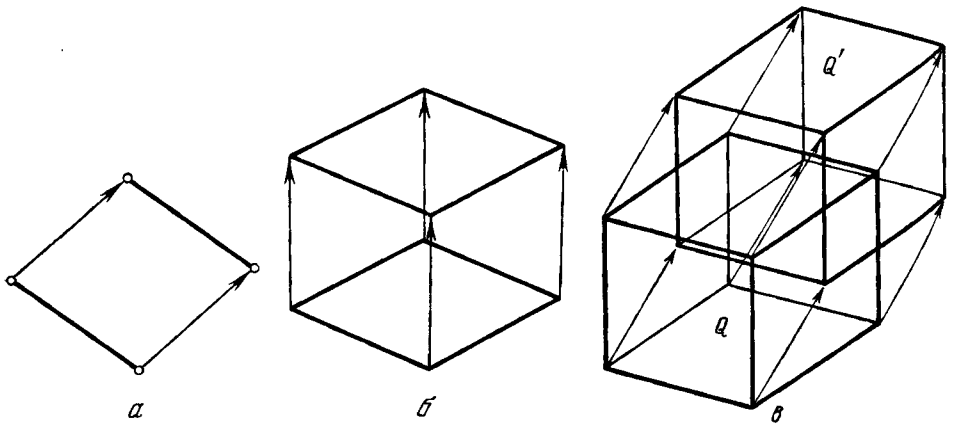


Рис. 23.

перемещении куба. Мы видим, что четырехмерный куб имеет всего 16 вершин: восемь у куба Q и восемь у куба Q'. Далее, он имеет 32 ребра: 12 ребер передвигаемого трехмерного куба в начальном положении Q, 12 ребер его в конечном положении Q' и 8 «боковых» ребер. Он имеет 8 трехмерных граней, которые сами являются кубами. При движении трехмерного куба каждая его грань зачерчивает трехмерный куб, так что получается 6 кубов — боковых граней четырехмерного куба, и, кроме того, имеются еще две грани: «передняя» и «задняя», соответственно первоначальному и конечному положению передвигаемого куба. Наконец, четырехмерный куб имеет еще двумерные квадратные грани общим числом 24: по шести у кубов Q и Q', и еще 12 квадратов, которые зачерчивают ребра куба Q при его перемещении.

Итак, четырехмерный куб имеет 8 трехмерных граней, 24 двумерных грани, 32 одномерных грани (32 ребра) и, наконец, 16 вершин; каждая грань есть «куб» соответствующего числа измерений: трехмерный куб, квадрат, отрезок, вершина (ее можно считать нульмерным кубом).

Аналогично, перемещая четырехмерный куб «в пятое измерение», получим пятимерный куб, и так, повторяя это построение, можно построить куб любого числа измерений. Все грани n-мерного куба сами

являются кубами меньшего числа измерений: $(n-1)$-мерные, $(n-2)$-мерные и т. д. и, наконец, одномерные, т. е. ребра. Любопытной и нетрудной задачей является найти, сколько граней каждого числа измерений имеет n-мерный куб. Легко убедиться, что он имеет $2n$ штук $(n-1)$-мерных граней и 2^n вершин. А сколько будет, например, ребер?

Рассмотрим еще один многогранник n-мерного пространства. На плоскости простейшим многоугольником является треугольник — он имеет наименьшее возможное число вершин. Чтобы получить многогранник с наименьшим числом вершин, достаточно взять точку, не лежащую в плоскости треугольника, и соединить ее отрезками с каждой точкой этого треугольника. Полученные отрезки заполнят трехгранную пирамиду —

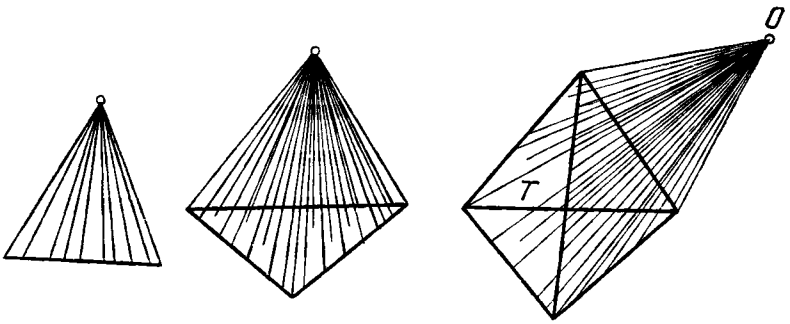


Рис. 24.

тетраэдр (рис. 24). Чтобы получить простейший многогранник в четырехмерном пространстве, рассуждаем так. Берем какую-нибудь трехмерную плоскость и в ней некоторый тетраэдр T. Затем, взяв точку, не лежащую в данной трехмерной плоскости, соединяем ее отрезками со всеми точками тетраэдра T. На самом правом из рис. 24 условно изображено это построение. Каждый из отрезков, соединяющих точку O с точкой тетраэдра T, не имеет с тетраэдром других общих точек, так как в противном случае он целиком помещался бы в трехмерном пространстве, содержащем T. Все такие отрезки как бы «идут в четвертое измерение». Они заполняют простейший четырехмерный многогранник — так называемый *четырехмерный симплекс*. Его трехмерные грани суть тетраэдры: один в основании и еще 4 боковых грани, опирающиеся на двумерные грани основания; всего 5 граней. Его двумерные грани — треугольники; их всего 10: четыре у основания и шесть боковых. Наконец, он имеет 10 ребер и 5 вершин.

Повторяя такое же построение для любого числа n измерений, получим простейший n-мерный многогранник — так называемый *n-мерный симплекс*. Как видно из построения, он имеет $n+1$ вершину. Можно убедиться, что все его грани тоже являются симплексами меньшего числа измерений: $(n-1)$-мерные, $(n-2)$-мерные и т. д.[1]

[1] Любые m вершин симплекса определяют «натянутый на них» $(m-1)$-мерный

Легко также обобщить понятия призмы и пирамиды. Если мы будем параллельно переносить многоугольник из плоскости в третье измерение, то он зачертит призму. Аналогично, перенося трехмерный многогранник в четвертое измерение, получим четырехмерную призму (условно это изображено на рис. 25). Четырехмерный куб есть, очевидно, частный случай призмы.

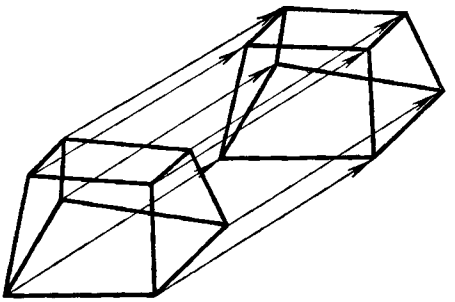


Рис. 25.

Пирамида строится следующим образом. Берется многоугольник Q и точка O, не лежащая в плоскости многоугольника. Каждая точка многоугольника Q соединяется отрезком с точкой O и эти отрезки заполняют пирамиду с основанием Q (рис. 26). Аналогично, если в четырехмерном пространстве дан трехмерный многогранник Q и точка O, не лежащая с ним в одной трехмерной плоскости, то отрезки, соединяющие точки многогранника Q с точкой O, заполняют четырехмерную пирамиду с основанием Q. Четырехмерный симплекс есть не что иное, как пирамида с тетраэдром в основании.

Совершенно аналогично, отправляясь от $(n-1)$-мерного многогранника Q, можно определить n-мерную призму и n-мерную пирамиду.

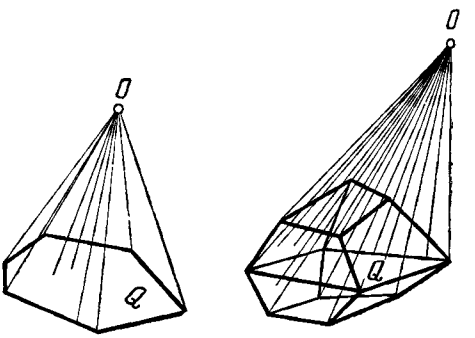


Рис. 26.

Вообще n-мерный многогранник есть часть n-мерного пространства, ограниченная конечным числом кусков $(n-1)$-мерных плоскостей; k-мерный многогранник есть часть k-мерной плоскости, ограниченная конечным числом кусков $(k-1)$-мерных плоскостей. Грани многогранника сами являются многогранниками меньшего числа измерений.

Теория n-мерных многогранников представляет собой богатое конкретным содержанием обобщение теории обычных трехмерных многогранников. В ряде случаев теоремы о трехмерных многогранниках обобщаются на любое число измерений без особого труда, но встречаются и такие

симплекс: $(m-1)$-мерную грань данного n-мерного симплекса. Число $(m-1)$-мерных граней n-мерного симплекса равно поэтому числу сочетаний из всех $n+1$ его вершин по m, т. е. равно

$$C_{n+1}^{m} = \frac{(n+1)!}{m!(n-m+1)!}.$$

вопросы, решение которых для n-мерных многогранников представляет огромные трудности. Здесь можно упомянуть глубокие исследования Г. Ф. Вороного (1868—1908), возникшие, кстати сказать, в связи с задачами теории чисел; они были продолжены советскими геометрами. Одна из возникших задач — так называемая «проблема Вороного» — все еще не решена полностью [1].

Примером, на котором обнаруживается существенная разница между пространствами разных измерений, могут служить правильные многогранники. На плоскости правильный многоугольник может иметь любое число сторон. Иными словами, имеется бесконечно много разных видов правильных «двумерных многогранников». Трехмерных правильных многогранников всего пять видов: тетраэдр, куб, октаэдр, додекаэдр, икосаэдр. В четырехмерном пространстве есть шесть видов правильных многогранников, но уже в любом пространстве большего числа измерений их всего три. Это: 1) аналог тетраэдра — правильный n-мерный симплекс, т. е. симплекс, все ребра которого равны; 2) n-мерный куб; 3) аналог октаэдра, который строится следующим образом: центры граней куба служат вершинами этого многогранника, так что он как бы натягивается на них. В случае трехмерного пространства это построение произведено на рис. 27. Мы видим, что в отношении правильных многогранников двух, трех и четырехмерные пространства занимают особое положение.

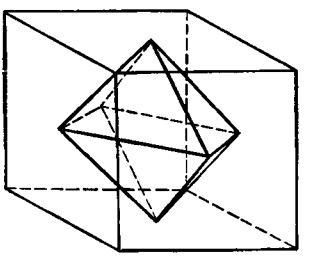


Рис. 27.

6. Рассмотрим еще вопрос об объеме тел в n-мерном пространстве. Объем n-мерного тела определяется аналогично тому, как это делается в обычной геометрии. Объем — это сопоставляемая фигуре численная характеристика, причем от объема требуется, чтобы у равных тел были равные объемы, т. е. чтобы объем не менялся при движении фигуры как твердого целого, и чтобы в случае, когда одно тело сложено из двух, его объем был равен сумме их объемов. За единицу объема принимается объем куба с ребром, равным единице. После этого устанавливается, что объем куба с ребром a равен a^n. Это делается так же, как на плоскости и в трехмерном пространстве, путем заполнения куба слоями из кубов (рис. 28). Так как кубы укладываются по n направлениям, то это и дает a^n.

Чтобы определить объем любого n-мерного тела, его заменяют приближенно телом, сложенным из весьма малых n-мерных кубов, подобно

[1] Речь идет об отыскании таких выпуклых многогранников, которыми можно заполнять пространство, прикладывая их друг к другу параллельно по целым граням. Для случая трехмерного пространства эта задача была поставлена и решена в связи с потребностями кристаллографии Федоровым; Вороной и его последователи продвинули ту же задачу для n-мерного пространства, но окончательно решена она только для пространств двух, трех и четырех измерений.

тому, как на рис. 29 плоская фигура заменяется фигурой из квадратиков. Объем тела определяется как предел таких ступенчатых тел при безграничном измельчении составляющих их кубиков.

Совершенно аналогично определяется k-мерный объем k-мерной фигуры, лежащей в какой-нибудь k-мерной плоскости. Из определения объема легко выводится важное его свойство: при подобном увеличении тела, когда все его линейные размеры увеличиваются в λ раз, k-мерный объем увеличивается в λ^k раз.

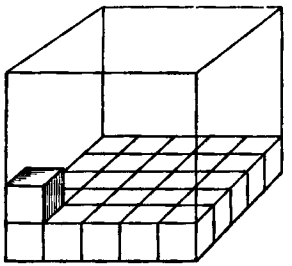


Рис. 28.

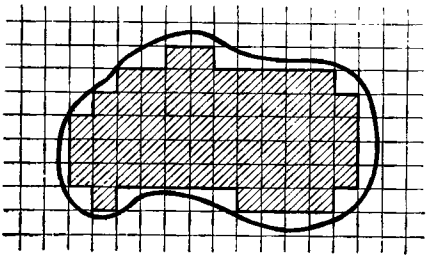


Рис. 29.

Если тело разбить на параллельные слои, то его объем представляется как сумма объемов этих слоев

$$V = \sum V_i.$$

Объем каждого слоя можно приближенно представить как произведение его высоты Δh_i на $(n-1)$-мерный объем («площадь») соответствующего сечения S_i. В результате объем всего тела приближенно представится суммой

$$V \approx \sum S_i \Delta h_i.$$

Переходя к пределу при всех $\Delta h_i \to 0$, получим представление объема в виде интеграла

$$V = \int_0^H S(h)\,dh, \tag{9}$$

где H — протяжение тела в направлении, перпендикулярном проводимым сечениям.

Все это совершенно аналогично вычислению объемов в трехмерном пространстве. Например, у призмы все сечения равны, и, следовательно, их «площадь» S не зависит от h. Поэтому для призмы $V = SH$, т. е. объем призмы равен произведению «площади» основания на высоту. Определим еще объем n-мерной пирамиды. Пусть дана пирамида с высотой H и площадью основания S. Пересечем ее плоскостью, параллельной основанию, на расстоянии h от вершины. Тогда отсечется пирамида высоты h. Площадь ее основания обозначим через $s(h)$. Эта мень-

шая пирамида, очевидно, подобна исходной: все ее размеры во столько же раз меньше, во сколько h меньше H, т. е. они умножаются на $\frac{h}{H}$. Поэтому $(n-1)$-мерный объем (т. е. «площадь») ее основания будет равен

$$s(h)=\left(\frac{h}{H}\right)^{n-1}S,$$

поскольку при изменении линейных размеров $(n-1)$-мерной фигуры в λ раз, ее объем умножается на λ^{n-1}.

Объем всей пирамиды согласно формуле (9) равен

$$V=\int_0^H s(h)\,dh,$$

откуда

$$V=\int_0^H S\left(\frac{h}{H}\right)^{n-1}dh=\frac{S}{H^{n-1}}\int_0^H h^{n-1}\,dh=\frac{S}{H^{n-1}}\cdot\frac{H^n}{n}=\frac{1}{n}SH,$$

т. е. объем n-мерной пирамиды равен $\frac{1}{n}$-й произведения «площади» [$(n-1)$-мерного объема] основания на высоту. При n, равном 2 или 3, получаем, как частные случаи, известные результаты: площадь (двумерный объем) треугольника равна половине произведения основания на высоту, а объем трехмерной пирамиды — одной трети произведения площади основания на высоту.

Шар можно приближенно представить как составленный из очень узких пирамид с общей вершиной в центре шара. Высоты этих пирамид равны радиусу R, а из площадей их оснований σ_i складывается приближенно вся поверхность шара S. Так как объем каждой пирамиды равен $\frac{1}{n}R\sigma_i$, то, складывая эти объемы, получим для объема шара

$$V\approx\frac{1}{n}R\sum\sigma_i\approx\frac{1}{n}RS.$$

В пределе это дает точную формулу: $V=\frac{1}{n}RS$, т. е. объем шара равен $\frac{1}{n}$-й произведения радиуса на его поверхность. Для n, равного 2 или 3, эта связь широко известна[1].

Отметим важное свойство шара, доказываемое для n-мерного пространства, вообще говоря, буквально так же, как для трехмерного: среди всех тел данного объема наименьшую площадь поверхности имеет шар и только шар.

[1] Вычисление объема шара можно осуществить также, применяя формулу (9); сечения n-мерного шара будут $(n-1)$-мерными шарами, и потому объем n-мерного шара вычисляется переходом от $n-1$ к n.

7. Мы ограничивались пока элементарной геометрией n-мерного пространства, но в нем можно развивать также «высшую» геометрию, например, общую теорию кривых и поверхностей. В n-мерном пространстве поверхности могут быть разного числа измерений: одномерные «поверхности», т. е. кривые, двумерные поверхности, трехмерные, ..., и, наконец, $(n-1)$-мерные поверхности. Кривую можно определить как геометрическое место точек, координаты которых непрерывно зависят от какой-либо переменной — параметра t

$$x_1 = x_1(t), \quad x_2 = x_2(t), \quad \ldots, \quad x_n = x_n(t).$$

Кривая есть как бы траектория движения точки в n-мерном пространстве с изменением t. Если наше пространство служит для изображения состояний какой-либо физической системы, как об этом говорилось в п. 4, то кривая изображает непрерывную последовательность состояний или ход изменения состояния в зависимости от параметра t (например, времени). Это обобщает обычное графическое изображение процесса изменения состояний посредством кривых.

С каждой точкой кривой в n-мерном пространстве связывают не только касательную («одномерную соприкасающуюся плоскость»), но и соприкасающиеся плоскости всех измерений от 2 до $n-1$. Скорость вращения каждой из этих плоскостей по отношению к скорости прохождения длины кривой дает соответствующую кривизну. Таким образом, кривая имеет $n-1$ соприкасающуюся плоскость, от одномерной до $(n-1)$-мерной, и соответственно $n-1$ кривизну. Дифференциальная геометрия в n-мерном пространстве оказывается гораздо более сложной, чем в трехмерном. На теории поверхностей мы не имеем возможности останавливаться.

До сих пор речь шла об n-мерной геометрии, обобщающей непосредственно обычную эвклидову геометрию. Но мы уже знаем, что, кроме эвклидовой геометрии, существует еще геометрия Лобачевского, проективная геометрия и др. Эти геометрии так же легко обобщаются на любое число измерений.

§ 8. ОБОБЩЕНИЕ ПРЕДМЕТА ГЕОМЕТРИИ

1. Говоря в предыдущем параграфе о реальном смысле n-мерного пространства, мы вплотную подошли к вопросу об обобщении предмета геометрии, к вопросу об общем понятии пространства в математике. Но прежде, чем дать соответствующие общие определения, рассмотрим еще ряд примеров.

Опыт показывает, что нормальное человеческое зрение, как это впервые отметил еще М. В. Ломоносов, трехцветно, т. е. всякое цветовое ощущение — цвет $Ц$ — есть комбинация трех основных ощущений:

красного $К$, зеленого $З$ и синего $С$, с определенными интенсивностями[1]. Обозначая эти интенсивности в некоторых единицах через x, y, z, можно написать, что $Ц = xК + yЗ + zС$. Подобно тому, как точку можно двигать в пространстве вверх и вниз, направо и налево, вперед и назад, так и ощущение цвета — цвет $Ц$ — может непрерывно изменяться в трех направлениях с изменением составляющих его частей красного, зеленого и синего. По аналогии можно поэтому сказать, что совокупность всевозможных цветов есть «трехмерное цветовое пространство». Интенсивности x, y, z играют роль координат точки — цвета $Ц$. (Важное отличие от обычных координат состоит в том, что интенсивности не могут быть отрицательными. Когда $x = y = z = 0$, получаем совершенно черный цвет, отвечающий полному отсутствию света.)

Непрерывное изменение цвета можно изображать как линию в «пространстве цветов»; такую линию образуют, например, цвета радуги; цветовую линию представляет также ряд ощущений, вызываемых предметом однородной окраски при непрерывном изменении яркости освещения. В этом случае меняется только интенсивность ощущения, его «цветность» остается неизменной.

Далее, если даны два цвета, скажем красный $К$ и белый $Б$, то, смешивая их в разных пропорциях[2], получим непрерывную последовательность цветов от $К$ до $Б$, которую можно назвать отрезком $КБ$. Представление о том, что розовый цвет лежит между красным и белым, имеет ясный смысл.

Таким образом, возникает понятие о простейших геометрических фигурах и отношениях в «пространстве цветов». «Точка» есть цвет, «отрезок» AB — совокупность цветов, получаемых смешением цветов A и B; то, что «точка D лежит на отрезке AB», означает, что цвет D есть смесь цветов A и B. Смешение трех цветов дает кусок плоскости — «цветовой треугольник». Все это можно также описать аналитически, пользуясь координатами цветов x, y, z, причем формулы, задающие цветовые прямые и плоскости, будут вполне аналогичны формулам обычной аналитической геометрии[3].

---

[1] Речь идет об ощущении цвета, а не о свете. Ощущение цвета есть также объективное явление — реакция на свет. Одно и то же ощущение может вызываться разными световыми волнами. Так, например, зеленый цвет может получаться не только от спектрально чистого зеленого света, но также и от смешения красного и синего. С другой стороны, у людей, страдающих «цветной слепотой» (дальтонизмом), есть только два основных ощущения; случаи «полной цветной слепоты», когда есть только одно основное ощущение цвета, крайне редки.

[2] Такое смешение можно получить, смешивая в разных пропорциях очень тонкие цветные порошки, при условии, что освещение остается неизменным.

[3] Например, если цвета $Ц_0$ и $Ц_1$ определяются интенсивностями — координатами x_0, y_0, z_0 и x_1, y_1, z_1, то цвет $Ц$, лежащий между $Ц_0$ и $Ц_1$, имеет ксординаты $x = (1-t)x_0 + tx_1$, $y = (1-t)y_0 + ty_1$, $z = (1-t)z_0 + tz_1$, где t есть доля цвета $Ц_1$, а $(1-t)$ — доля цвета $Ц_0$ в смеси, образующей цвет $Ц$.

В пространстве цветов выполняются соотношения эвклидовой геометрии, касающиеся расположения точек и отрезков. Учение об этих отношениях образует аффинную геометрию, и можно сказать, что в совокупности всех возможных цветовых ощущений реализуется аффинная геометрия. (Это, правда, не вполне точно, потому что, как уже сказано, координаты цвета x, y, z не могут быть отрицательными. Поэтому пространство цветов отвечает только той части пространства, где в данной системе координат все координаты точек положительны или нули.)

Далее, мы имеем естественное представление о степени различия цветов. Так, например, понятно, что бледнорозовый ближе к белому, чем густой розовый, а малиновый ближе к красному, чем синий, и т. д. Таким образом, мы имеем качественное понятие о расстоянии между цветами, как о степени их различия. Этому качественному понятию можно дать количественную меру. Однако определять расстояние между цветами так же, как в эвклидовой геометрии по формуле $r=\sqrt{(x_0-x_1)^2+(y_0-y_1)^2+(z_0-z_1)^2}$, оказывается неестественным. Так определяемое расстояние не соответствует реальному ощущению; при таком определении получалось бы в ряде случаев, что два цвета, в разной степени отличные от данного, находились бы от него на одном расстоянии. Определение расстояния должно отображать реальные отношения между цветовыми ощущениями.

Руководствуясь этим, в пространстве цветов вводят особое измерение расстояний. Делают это следующим образом.

При непрерывном изменении цвета человек не сразу ощущает это изменение, а лишь тогда, когда оно достигает известной степени, доходя до так называемого порога различения. В связи с этим считают, что все цвета, находящиеся от данного как раз на пороге различения, равноудалены от него. Тогда, само собой, приходим к тому, что расстояние между любыми двумя цветами нужно измерять наименьшим числом порогов различения, которое только можно проложить между ними. Длина цветовой линии измеряется числом укладывающихся на ней таких порогов. Расстояние между двумя цветами определяется длиной самой короткой соединяющей их линии. Это сходно с тем, как расстояние двух точек на поверхности измеряют длиной самой короткой соединяющей их линии.

Таким образом, измерение длин и расстояний в цветовом пространстве производится очень малыми, как бы бесконечно малыми шагами.

В результате в пространстве цветов определяется некоторая своеобразная неэвклидова геометрия. Эта геометрия имеет вполне реальный смысл: она описывает на геометрическом языке свойства совокупности всевозможных цветов, т. е. свойства реакций глаза на световое раздражение.

Понятие о цветовом пространстве возникло около ста лет назад. Геометрию этого пространства изучали многие физики, из которых можно назвать, например, Гельмгольца и Максвелла. Эти исследования продол-

жаются; они имеют не только теоретическое, но и практическое значение. Они дают точную математическую основу для решения вопросов о различении цвета сигналов, о красках в текстильной промышленности и др.

2. Рассмотрим другой пример, о котором уже говорилось в предыдущем параграфе.

Пусть мы изучаем какую-либо физико-химическую систему, как то: смесь газов, сплав и т. п. Пусть состояние этой системы определяется n величинами (так, состояние газовой смеси определяется давлением, температурой и концентрациями составляющих ее компонентов). Тогда говорят, что система имеет n степеней свободы, выражая этим, что ее состояние может меняться, так сказать, в n независимых направлениях с изменением каждой из определяющих это состояние величин. Эти величины, определяющие состояние системы, играют как бы роль его координат. Поэтому совокупность всех ее состояний рассматривают как n-мерное пространство — так называемое фазовое пространство системы.

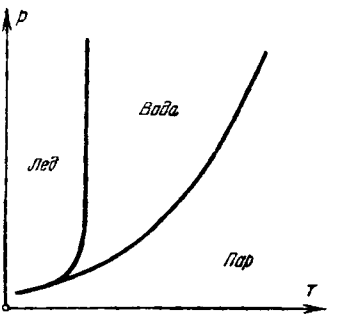


Рис. 30.

Непрерывные изменения состояния, т. е. процессы, происходящие в системе, изображаются линиями в этом пространстве. Отдельные области состояний, выделяемые по тем или иным признакам, будут областями фазового пространства. Состояния, пограничные для двух областей, образуют поверхности в этом пространстве.

В физической химии особенно важно изучить форму и взаимное прилегание тех областей фазового пространства системы, которые соответствуют качественно различным состояниям. Поверхности, разделяющие эти области, соответствуют таким качественным переходам, как плавление, испарение, выпадение осадка и т. п. Состояние системы с двумя степенями свободы изображается точкой на плоскости. Примером может служить однородное вещество, состояние которого определяется давлением p и температурой T; они и служат координатами точки, изображающей состояние. Тогда дело сводится к линиям раздела областей, отвечающих качественно различным состояниям. В случае воды, например, такими областями будут области льда, жидкой воды и пара (рис. 30). Линии их раздела отвечают плавлению (затвердеванию), испарению (конденсации), возгонке льда (выпадению кристаллов льда из пара).

При изучении систем со многими степенями свободы требуются методы многомерной геометрии.

Понятие фазового пространства применяется не только к физико-химическим, но также к механическим системам, и вообще оно может применяться к любой системе, если ее возможные состояния образуют некоторую непрерывную совокупность. В кинетической теории газов

рассматривают, например, фазовые пространства системы материальных частиц — молекул газа. Состояние движения одной частицы в каждый момент определяется ее положением и скоростью, что дает всего шесть величин: три координаты и три составляющие скорости (по трем осям координат). Состояние N частиц задается $6N$ величинами, и так как молекул очень много, то $6N$ — огромное число. Это нимало не смущает физиков, говорящих о $6N$-мерном фазовом пространстве системы молекул.

Точка в этом пространстве изображает состояние всей массы молекул с их координатами и скоростями. Движение точки изображает изменение состояния. Такое абстрактное представление оказывается очень полезным во многих глубоких теоретических выводах. Словом, понятие фазового пространства прочно вошло в арсенал точного естествознания и применяется в разнообразных вопросах.

3. Приведенные примеры позволяют уже придти к выводу о том, как обобщается предмет геометрии.

Пусть мы исследуем какую-либо непрерывную совокупность тех или иных объектов, явлений или состояний, например совокупность всевозможных цветов или совокупность состояний группы молекул. Отношения, имеющиеся в такой совокупности, могут оказаться сходными с обычными пространственными отношениями, как «расстояние» между цветами или «взаимное расположение» областей фазового пространства. В таком случае, отвлекаясь от качественных особенностей изучаемых объектов и принимая во внимание только эти отношения между ними, мы можем рассматривать данную совокупность как своего рода пространство. «Точками» этого «пространства» служат сами объекты, явления или состояния. «Фигурой» в таком пространстве будет любая совокупность его точек, как, например, «линия» цветов радуги или «область» пара в «пространстве» состояний воды. «Геометрия» такого пространства определяется как раз теми пространственно-подобными отношениями, которые имеются между данными объектами, явлениями или состояниями. Так, «геометрия» цветового пространства определяется законами смешения цветов и расстояниями между цветами.

Реальное значение такой точки зрения состоит в том, что она дает возможность использовать понятия и методы отвлеченной геометрии для изучения разнообразнейших явлений. Область применения геометрических понятий и методов расширяется, таким образом, чрезвычайно. В результате обобщения понятия пространства термин «пространство» получает в науке два значения: с одной стороны, это обычное реальное пространство — универсальная форма существования материи, с другой стороны, это «абстрактное пространство» — совокупность однородных объектов (явлений, состояний и т. п.), в которой имеются пространственно-подобные отношения.

Стоит заметить, что обычное пространство, как мы его себе несколько упрощенно представляем, тоже можно понимать как совокуп

ность однородных состояний. Именно, его можно понимать как совокупность всех возможных положений предельно малого тела — «материальной точки». Это замечание не претендует на то, чтобы дать определение пространства, но имеет целью сделать более ясной связь двух понятий пространства. Понятие абстрактного пространства мы еще поясним ниже, об отношении же абстрактной геометрии к обычному реальному пространству речь будет идти в последнем параграфе этой главы.

4. Наиболее широкое применение понятие абстрактного пространства находит в самой математике. В геометрии рассматривают «пространство» тех или иных фигур, как, например, «пространство шаров», о котором уже шла речь, «пространство прямых» и т. п.

Этот метод оказывается, в частности, чрезвычайно плодотворным в теории многогранников. Так, например, в § 5 главы VII (том 2) упоминалась теорема о существовании выпуклого многогранника с данной разверткой. Доказательство этой теоремы основано на рассмотрении двух «пространств»: «пространства многогранников» и «пространства разверток». Совокупность выпуклых многогранников, имеющих данное число вершин, рассматривается как своего рода пространство, где точка изображает многогранник; соответственно совокупность допустимых разверток также трактуется как некоторое пространство, где точка изображает развертку. Склеивание многогранников из разверток устанавливает соответствие между многогранниками и развертками, т. е. соответствие между точками «пространства многогранников» и «пространства разверток». Задача состоит в том, чтобы доказать, что каждой развертке отвечает многогранник, т. е. каждой точке одного пространства отвечает точка другого. Это как раз и доказывается посредством применения топологии.

Аналогично доказывается целый ряд других теорем о многогранниках, причем этот «метод абстрактных пространств» в ряде случаев (как в теореме о существовании многогранника с данной разверткой) оказывается самым простым из известных методов доказательства таких теорем. К сожалению, однако, этот метод все же довольно сложен, и мы не имеем здесь возможности дать о нем более точное представление.

Широкое применение обобщенное понятие пространства находит также в анализе, алгебре и теории чисел. Начало этого лежит в обычном представлении функций посредством кривых. Значениям одной переменной x обычно сопоставляются точки на прямой. Аналогично, значениям двух переменных сопоставляются точки на плоскости, значениям n переменных — точки в n-мерном пространстве; набор значений переменных $x_1, x_2, \ldots, x_n$ представляют точкой с координатами $x_1, x_2, \ldots, x_n$. Говорят об «области изменения переменных» или об «области задания» функции этих переменных $f(x_1, x_2, \ldots, x_n)$; говорят о точках, линиях или поверхностях разрыва функции и т. п. Этот геометрический язык употребляется постоянно, и это не только способ выражения; геометрические представ-

ления делают многие факты анализа «наглядными» по аналогии с обычным пространством и позволяют применять геометрические методы доказательства, обобщенные на n-мерное пространство.

То же имеет место в алгебре, когда речь идет об уравнениях с n-неизвестными или об алгебраических функциях от n переменных. В предыдущем параграфе отмечалось, что линейное уравнение с n неизвестными определяет в n-мерном пространстве плоскость, m таких уравнений определяют m плоскостей, а всякое их решение изображает точку, общую для всех этих плоскостей. Плоскости могут вовсе не пересекаться, пересекаться в одной точке, по целой прямой, по двумерной или вообще по некоторой k-мерной плоскости. В целом вопрос о разрешимости системы линейных уравнений изображается как вопрос о пересечении плоскостей. Этот геометрический подход имеет ряд преимуществ. Вообще, «линейную алгебру», которая объемлет учение о линейных уравнениях и линейных преобразованиях, обычно излагают в большой степени геометрически, как это сделано в главе XVI.

5. Во всех рассмотренных примерах речь шла о том, что непрерывная совокупность тех или иных объектов трактуется как своего рода пространство. Объектами этими служили цвета, состояния той или иной системы, фигуры, совокупности значений переменных. Во всех случаях объект задавался конечным числом данных, и потому соответствующее пространство имело конечное число измерений, равное числу этих данных.

Однако в начале нашего столетия в математике начали рассматривать также «бесконечномерные пространства» — совокупности таких объектов, каждый из которых не может быть задан конечным числом данных. Это прежде всего «функциональные пространства».

Мысль трактовать совокупность функций того или иного типа как своего рода пространство является одной из основных идей новой отрасли анализа — функционального анализа — и оказывается чрезвычайно плодотворной в решении многих вопросов. Читатель получит об этом представление из главы XIX, специально посвященной функциональному анализу.

Можно рассматривать пространства непрерывных функций одной или нескольких переменных. Как «пространства» рассматривают также различные классы разрывных функций, определяя расстояние между функциями тем или иным способом в зависимости от характера подлежащих решению задач. Словом, число возможных «функциональных пространств» неограничено, и фактически в математике изучают много таких пространств.

Совершенно так же можно рассматривать «пространство кривых», «пространство выпуклых тел», «пространство возможных движений данной механической системы» и т. п. В § 5 главы VII (том 2) были, например, упомянуты теоремы о том, что на всякой замкнутой выпуклой по-

верхности существуют по крайней мере три замкнутые геодезические и что каждые две точки соединимы бесконечным числом геодезических линий. При доказательствах этих теорем используются пространства кривых на поверхности: в первой теореме — пространство замкнутых кривых, во второй — пространство кривых, соединяющих две данные точки. В совокупности всевозможных кривых, соединяющих две данные точки, вводится своего рода расстояние, и таким путем эта совокупность превращается в пространство. Доказательство теоремы основывается на применении некоторых глубоких результатов топологии к этому пространству.

Формулируем теперь общий вывод.

Под «пространством» в математике понимают вообще любую совокупность однородных объектов (явлений, состояний, функций, фигур, значений переменных и т. п.), между которыми имеются отношения, подобные обычным пространственным отношениям (непрерывность, расстояние и т. п.). При этом, рассматривая данную совокупность объектов как пространство, отвлекаются от всех свойств этих объектов, кроме тех, которые определяются этими принятыми во внимание пространственно-подобными отношениями. Эти отношения определяют то, что можно назвать строением или «геометрией» пространства. Сами объекты играют роль «точек» такого пространства; «фигуры» — это множества его «точек».

Предмет геометрии данного абстрактного пространства составляют те свойства пространства и фигур в нем, которые определяются принятыми в расчет пространственно-подобными отношениями. Так, например, при рассмотрении пространства непрерывных функций вовсе не занимаются свойствами отдельных функций самих по себе. Функция играет здесь роль точки и, стало быть, «не имеет частей», не имеет в этом смысле никакого строения, никаких свойств вне связи с другими точками; точнее, от всего этого отвлекаются. В функциональном пространстве свойства функций определяют только через их отношения друг к другу — через их расстояния и через другие отношения, которые можно вывести из расстояния.

Разнообразию возможных совокупностей объектов и различных отношений между ними отвечает неограниченное разнообразие пространств, изучаемых в математике. Пространства можно классифицировать по типам тех пространственно-подобных отношений, которые кладутся в основу их определения. Не имея в виду охватить все разнообразие типов абстрактных пространств, отметим прежде всего два наиболее важных типа: топологические и метрические пространства.

6. *Топологическим пространством* (см. главу XVIII) называют любую совокупность — множество каких-либо элементов — точек, где определено отношение прикосновения одной точки к множеству точек и соответственно отношение прикосновения или прилегания двух множеств (фигур) друг к другу. Это есть обобщение наглядно понятного отношения прикосновения или прилегания фигур в обычном пространстве.

Еще Лобачевский с замечательной прозорливостью указывал, что из всех отношений фигур самым основным является отношение прикосновения. «*Прикосновение* составляет отличительную принадлежность тел и дает им название *геометрических*, когда в них удерживаем это свойство, не принимая в рассуждение все другие, существенные ли то будут или случайные»[1]. Например, любая точка на окружности прилегает к совокупности всех внутренних точек круга; две части связного целого тела прилегают друг к другу. Как показало дальнейшее развитие топологии, именно свойства прикосновения лежат в основе всех остальных топологических свойств.

Понятие прикосновения выражает представление о бесконечной близости точки к множеству. Поэтому всякая совокупность объектов, где имеется естественное понятие о непрерывности, о бесконечном приближении, тем самым оказывается топологическим пространством.

Понятие топологического пространства является чрезвычайно общим, и учение о таких пространствах — абстрактная топология — представляет собой не что иное, как наиболее общее математическое учение о непрерывности.

Строго математическое определение общего топологического пространства может быть дано следующим образом.

Общим топологическим пространством называем множество R каких-либо элементов — «точек», если в этом множестве для каждого содержащегося в нем множества M определены точки прикосновения так, что выполнены следующие условия — аксиомы:

1) Каждая точка множества M считается его точкой прикосновения. (Вполне естественно считать, что множество прикасается к каждой своей точке.)

2) Если множество M_1 содержит множество M_2, то множество точек прикосновения M_1 заведомо содержит все точки прикосновения M_2. (Говоря короче, но менее точно, у большего множества не меньше точек прикосновения.)

К этим аксиомам добавляют обычно еще другие, определяя таким путем те или иные типы топологических пространств.

Пользуясь понятием прикосновения, легко определить ряд важнейших топологических понятий. Эти понятия являются вместе с тем наиболее основными и общими понятиями геометрии, а их определения оказываются наглядно вполне ясными. Приведем примеры.

1) *Прилегание множеств.* Будем говорить, что множества M_1 и M_2 прилегают друг к другу, если одно из них содержит хотя бы одну точку прикосновения другого. (В этом смысле, например, окружность прилегает к внутренности круга.)

2) Непрерывность, или, как говорят математики, *связность* фигуры. Фигура, т. е. множество точек M, связна, если ее нельзя разбить на неприлегающие друг к другу части. (Например, отрезок связен, а отрезок без средней точки несвязен.)

---

[1] Н. И. Лобачевский, Собрание сочинений, т. II, 1949, стр. 168.

3) *Граница.* Границей множества M в пространстве R называется множество всех точек, которые прилегают как к самому M, так и к его дополнению $R - M$, т. е. к остальной части пространства R. (Это, очевидно, вполне естественное понятие границы.)

4) *Внутренняя точка.* Точка множества M называется внутренней, если она не лежит на его границе, т. е. если она не прилегает к его дополнению $R - M$.

5) *Непрерывное отображение или преобразование.* Преобразование множества M называется непрерывным, если оно не нарушает прикосновений. (Едва ли можно дать более естественное определение непрерывного преобразования.)

К этому списку можно было бы добавить еще другие важные определения, как, например, определение понятия сходимости последовательности фигур к данной фигуре или понятие о числе измерений пространства.

Мы видим, что через прикосновение определяются наиболее основные геометрические понятия. Значение топологии, в частности, состоит в том, что она дает этим понятиям строгие общие определения, дает основание для строгого применения соображений, связанных с наглядно понимаемой непрерывностью.

Топология есть учение о тех свойствах пространств, фигур в них и их преобразований, которые определяются отношением прикосновения.

Общность и фундаментальность этого отношения делает топологию наиболее общей геометрической теорией, проникающей в различные области математики — всюду, где только речь идет о непрерывности. Но точно так же в силу своей общности топология в своих наиболее абстрактных отделах уже выходит за рамки собственно геометрии. И все-таки в основе ее лежит обобщение свойств реального пространства, и наиболее плодотворные и сильные ее результаты связаны с применением методов, имеющих источник в наглядных геометрических представлениях. Таков, например, метод приближения общих фигур многогранниками, развитый П. С. Александровым и распространенный им, хотя и в абстрактной форме, на чрезвычайно общие типы топологических пространств.

Сейчас любой специалист, какие бы объекты он ни исследовал, обнаружив, что для них естественным образом вводится понятие близости, прилегания, немедленно получает в свои руки уже готовый, тонко разветвленный аппарат топологии, позволяющий делать выводы, далеко не тривиальные даже в их частных проявлениях.

7. *Метрическим пространством* называется множество каких-либо элементов — точек, между которыми определено расстояние, т. е. каждой паре точек X, Y отнесено число $r(X, Y)$ так, что выполнены следующие условия — аксиомы метрического пространства:

1) $r(X, Y) = 0$ тогда и только тогда, когда точки X, Y совпадают.

2) Для любых трех точек X, Y, Z

$$r(X, Y)+r(Y, Z)\geqslant r(Z, X).$$

Это условие называется «неравенством треугольника», поскольку оно вполне аналогично известному свойству обычного расстояния между точками A_1, A_2, A_3 эвклидова пространства (рис. 31):

$$r(A_1, A_2)+r(A_2, A_3)\geqslant r(A_3, A_1).$$

Примерами метрических пространств могут служить:

1) эвклидово пространство любого числа измерений n,

2) пространство Лобачевского,

3) любая поверхность с ее внутренней метрикой (том 2, глава VII, § 4),

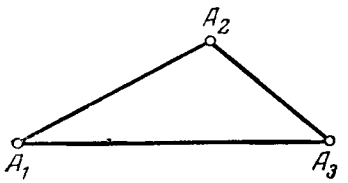


Рис. 31.

4) пространство C непрерывных функций с расстоянием, определенным по формуле

$$r(f_1, f_2)=\max|f_1(x)-f_2(x)|,$$

5) описываемое в главе XIX гильбертово пространство, которое есть не что иное, как «бесконечномерное эвклидово» пространство.

Гильбертово пространство является важнейшим из пространств, применяемых в функциональном анализе; оно тесным образом связано с теорией рядов Фурье и вообще с теорией разложения функций в ряды по ортогональным функциям (координаты x_1, x_2, x_3, ... оказываются здесь коэффициентами таких рядов). Это пространство играет важную роль в математической физике и приобрело большое значение в квантовой механике. Оказывается, что совокупность всевозможных (не только стационарных) состояний какой-либо атомной системы, например атома водорода, с абстрактной точки зрения может рассматриваться как гильбертово пространство.

Число примеров метрических пространств, фактически рассматриваемых в математике, можно было бы еще значительно умножить; кстати, в следующем параграфе мы познакомимся с одним важным классом метрических пространств, называемых римановыми, но и приведенных пока примеров достаточно, чтобы видеть широту объема общего понятия метрического пространства.

В метрическом пространстве всегда можно определить все топологические понятия, а сверх того ввести также другие — «метрические» понятия. Таково, например, понятие длины кривой. Длина определяется в любом метрическом пространстве буквально так же, как обычно, и основные ее свойства при этом сохраняются. Именно, под длиной кривой понимают предел сумм расстояний между точками, последовательно расположенными на кривой X_1, X_2, ..., X_n, при условии, что эти точки располагаются на кривой все гуще и гуще.

8. В математике рассматривают многие типы пространств помимо общих топологических или метрических. С целым классом таких пространств мы уже, собственно говоря, познакомились в § 6. Это пространства, в каждом из которых задана какая-либо группа преобразований (например, пространства проективное и аффинное). В таких пространствах можно определить «равенство» фигур. Фигуры «равны», если они переводятся одна в другую преобразованием из данной группы.

Мы не будем углубляться в определения возможных типов пространств; они достаточно разнообразны, и читатель может обратиться к специальной литературе по различным современным разделам геометрии.

Какой, однако, смысл в том, чтобы так расширять круг геометрических понятий? Для чего, например, нужно вводить понятие о пространстве непрерывных функций? Не достаточно ли решать задачи анализа его обычными средствами, не поднимаясь до таких абстрактных построений?

Общий ответ на этот вопрос состоит коротко в том, что, вводя в рассмотрение то или иное пространство, мы открываем путь применению геометрических понятий и методов, которые могут дать очень много.

Особенность геометрических понятий и методов состоит в том, что они основаны в конечном счете на наглядных представлениях и сохраняют, хотя бы и в абстрактной форме, их преимущества. То, чего аналитик достигает путем долгих выкладок, геометр может порой ухватить сразу. Начальный пример тому можно видеть в графике, который делает совершенно ясным ход той или иной зависимости между величинами. Геометрический метод можно характеризовать как метод синтетический, охватывающий целое, в отличие от метода аналитического. Конечно, в абстрактных геометрических теориях непосредственная наглядность исчезает, но остаются наглядные соображения по аналогии, остается синтетический характер геометрического метода.

Читатель уже знаком с применением геометрических картин в анализе, с геометрическим изображением комплексных чисел и функций от комплексной переменной, с геометрическими рассуждениями в доказательстве основной теоремы алгебры и с другими применениями геометрических понятий и методов. Всюду он может заметить то, о чем мы говорим здесь в общем виде. Мы упоминали в начале § 7, а также здесь, в п. 5, примеры теорем, доказываемых посредством применения многомерной геометрии. Укажем еще один пример задачи из анализа, которая решается уже на основе применения понятия о функциональном пространстве.

В топологии доказывается, что если взять на обычной плоскости любую область, имеющую форму деформированного круга, и затем как

угодно деформировать ее непрерывным образом, так, однако, чтобы в итоге она оказалась вложенной внутрь своего первоначального контура, то хотя бы одна точка области попадает после преобразования на место, где она была раньше. Этот факт — чисто топологический.

Рассмотрим теперь совершенно далекую от геометрии проблему: разыскивается функция $y(x)$, удовлетворяющая дифференциальному уравнению[1]

$$y' = f(x, y) \tag{10}$$

и принимающая при $x = 0$ значение $y = 0$.

Очевидно, вместо этого уравнения можно искать решение уравнения

$$y = \int_0^x f(t, y(t))\, dt. \tag{11}$$

Возникает естественный вопрос, а существует ли вообще удовлетворяющая этому условию функция $y(x)$?

Взглянем на задачу иначе. Представим себе всякую непрерывную функцию $y(x)$ точкой некоторого абстрактного пространства. Результат вычисления интеграла

$$\int_0^x f(t, y(t))\, dt = z(x)$$

будет снова непрерывной функцией от x, т. е. «точкой» нашего абстрактного пространства. Беря различные «точки» y, т. е. разные функции $y(x)$, будем получать, вообще говоря, разные точки z. Тем самым совокупность точек нашего пространства отображается снова в его же точки. Вопрос о наличии решения уравнения (11) свелся к вопросу о том, найдется ли «точка» нашего абстрактного пространства, которая после такого преобразования попадет на свое прежнее «место»?

Естественный вопрос из теории дифференциальных уравнений оказался вопросом о свойстве абстрактного функционального пространства. Аналогия с упомянутой выше теоремой подсказывает нам, что речь идет, повидимому, о топологическом свойстве соответствующего пространства.

На этом пути при помощи необходимых топологических исследований получают, пожалуй, наиболее краткие доказательства многих теорем о существовании решений дифференциальных уравнений, в частности выясняют, что уравнение (10) действительно имеет решение при любой непрерывной функции $f(x, y)$.

---

[1] $f(x, y)$ предполагается непрерывной функцией переменных x и y.

§ 9. РИМАНОВА ГЕОМЕТРИЯ

1. Изложенная выше идея о том, что любую непрерывную совокупность однородных явлений можно трактовать как своего рода пространство, была впервые высказана Риманом в его лекции «О гипотезах, лежащих в основании геометрии», прочитанной в Геттингенском университете в 1854 г. Это была пробная лекция, нечто вроде доклада или диссертации, которую поступающий на должность доцента или профессора должен был читать перед факультетом. В своей лекции Риман наметил в общих чертах без выкладок и математических доказательств исходные идеи обширной геометрической теории, называемой теперь римановой геометрией. Говорят, что никто из слушателей ее не понял, кроме старого уже тогда Гаусса. Формальный аппарат теории Риман изложил в другой работе в применении к задаче теплопроводности, так что абстрактная риманова геометрия рождалась в тесной связи с математической физикой. Идеи Римана были следующим после Лобачевского решающим шагом в развитии геометрии. Однако работы Римана не были сразу оценены должным образом. Его лекция и работа о теплопроводности были опубликованы лишь в 1868 г., после его смерти. Стоит отметить, что в 1868 г. появилось первое истолкование геометрии Лобачевского, данное Бельтрами, а в 1870 г. — ее второе истолкование, данное Клейном. В 1872 г. Клейн формулирует общий взгляд на различные геометрии: эвклидову, Лобачевского, проективную, аффинную и др., как на учение о свойствах фигур, не изменяющихся при преобразованиях из той или иной группы. В те же годы окончательно укрепляется в математике многомерная геометрия. Таким образом, 70-е годы XIX в. были тем переломным моментом в истории геометрии, когда новые геометрические идеи, складывавшиеся в течение предшествующих пятидесяти лет, наконец были поняты широким кругом математиков и прочно вошли в науку. Тогда работы Римана были продолжены, а к концу прошлого столетия риманова геометрия уже достигла значительного развития и нашла приложения в механике и физике. Когда же в 1915 г. Эйнштейн в своей общей теории относительности применил риманову геометрию к теории всемирного тяготения, это привлекло к римановой геометрии особое внимание, и в результате последовало ее бурное развитие и разнообразные обобщения.

2. Идеи Римана, имеющие столь блестящее продолжение, оказываются достаточно простыми, если отвлечься от математической разработки и обратить внимание лишь на их основную сущность. Такая простота свойственна собственно всем большим идеям. Не была ли проста идея Лобачевского, рассматривать выводы из отрицания V постулата как некоторую возможную геометрию? Не проста ли идея эволюции организмов или идея атомного строения вещества? Все это просто и вместе с тем очень сложно, ибо новые идеи, во-первых,

должны прокладывать себе широкую дорогу, а не укладываться в сложившиеся рамки, а во-вторых, их многостороннее обоснование, развитие и применение требуют массы труда, изобретательности и невозможны без специального аппарата науки. Для римановой геометрии таким аппаратом служат ее формулы; они сложны и потому доступны лишь специалистам. Но мы не будем думать о сложных формулах, а обратимся к сущности идей Римана. Как уже сказано, Риман начинает с рассмотрения любой непрерывной совокупности явлений как своего рода пространства. В этом пространстве координатами точки являются величины, определяющие соответствующее явление среди других, как, например, интенсивности x, y, z, определяющие цвет $Ц = xК + yЗ + zС$. Если таких величин n, скажем $x_1, x_2, \ldots, x_n$, то говорим об n-мерном пространстве. В этом пространстве можно рассматривать линии и ввести измерение их длин малыми (бесконечно малыми) шагами, подобно измерению длины кривой в обычном пространстве.

Для того чтобы мерить длины бесконечно малыми шагами, достаточно задать закон, определяющий расстояние от любой данной точки до любой к ней бесконечно близкой. Этот закон определения (измерения) расстояний называют *мероопределением* или *метрикой*. Самый простой случай, — когда этот закон оказывается тем же, что в эвклидовом пространстве. Такое пространство будет эвклидовым в бесконечно малом. Иными словами, геометрические соотношения эвклидовой геометрии будут в нем выполняться, но лишь в каждой бесконечно малой области; правильнее сказать, они выполняются в любой достаточно малой области, но не точно, а с точностью тем большей, чем меньше область. Пространство, в котором расстояния измеряются по такому закону, называют *римановым*; геометрию же таких пространств называют *римановой*. Риманово пространство есть, следовательно, пространство, являющееся эвклидовым «в бесконечно малом».

Простейший пример риманова пространства представляет любая гладкая поверхность с ее внутренней геометрией. Внутренняя геометрия поверхности есть риманова геометрия двух измерений. Действительно, вблизи каждой своей точки гладкая поверхность мало отличается от касательной плоскости, и это отличие тем меньше, чем меньшую область поверхности мы рассматриваем. Поэтому и геометрия в малой области поверхности мало отличается от геометрии на плоскости; чем область меньше, тем меньше это отличие. Однако в больших областях геометрия кривой поверхности оказывается отличной от эвклидовой, как это было выяснено в § 4 главы VII (том 2) и как это легко видеть на примерах шара или псевдосферы. Риманова геометрия есть не что иное, как естественное обобщение внутренней геометрии поверхности с двух измерений на любое число n. Подобно поверхности, рассматриваемой лишь с точки зрения ее внутренней геометрии, трехмерное риманово пространство, будучи эвкли-

довым в малых областях, в больших областях может отличаться от эвклидова. Например, длина окружности может не быть пропорциональной радиусу; она будет с хорошим приближением пропорциональной радиусу лишь для малых окружностей. Сумма углов треугольника может не равняться двум прямым (при этом в построении треугольника роль прямолинейных отрезков играют линии кратчайшего расстояния, т. е. линии, имеющие наименьшую длину среди всех линий, соединяющих данные точки).

Можно представить себе, что реальное пространство оказывается эвклидовым лишь в областях, небольших в сравнении с астрономическими масштабами. Чем меньше область, тем точнее выполняется эвклидова геометрия, но можно думать (и так оно и оказывается), что в очень больших масштабах геометрия уже несколько отлична от эвклидовой. Эту идею, как мы знаем, выдвинул еще Лобачевский. Риман обобщил ее, допуская мысль о любой геометрии, эвклидовой в бесконечно малом, а не только о геометрии Лобачевского, которая оказывается частным случаем римановой геометрии.

Из сказанного видно, что риманова геометрия возникла на основе синтеза и обобщения трех идей, выдвинутых развитием геометрии. Первой явилась идея о возможности геометрии, отличной от эвклидовой; второй было понятие о внутренней геометрии поверхностей третьей — понятие о пространстве любого числа измерений.

3. Для того чтобы уяснить, как математически определяется риманово пространство, вспомним прежде всего закон измерения расстояний в эвклидовом пространстве.

Если на плоскости ввести прямоугольные координаты x, y, то по теореме Пифагора расстояние между двумя точками, координаты которых отличаются на Δx и Δy, выражается формулой

$$s = \sqrt{\Delta x^2 + \Delta y^2}.$$

В трехмерном пространстве аналогично

$$s = \sqrt{\Delta x^2 + \Delta y^2 + \Delta z^2}.$$

В n-мерном же эвклидовом пространстве расстояние определяют общей формулой

$$s = \sqrt{\Delta x_1^2 + \ldots + \Delta x_n^2}.$$

Отсюда легко заключить, как нужно задавать закон измерения расстояний в римановом пространстве. Закон этот должен совпадать с эвклидовским, но лишь в бесконечно малой области вблизи каждой точки. Это приводит к следующей формулировке этого закона.

Риманово n-мерное пространство характеризуется тем, что вблизи любой его точки A можно ввести координаты $x_1, x_2, \ldots, x_n$ так, чтобы расстояние от точки A до бесконечно близкой точки X выражалось формулой

$$ds = \sqrt{dx_1^2 + \ldots + dx_n^2}, \qquad (12)$$

где $dx_1, \ldots, dx_n$ — бесконечно малые разности координат точек A и X. Точнее это можно выразить иначе, а именно: расстояние от точки A до любой близкой точки X выражается той же формулой, что в эвклидовой геометрии, но лишь с некоторой точностью, которая тем выше, чем ближе точка X к точке A, т. е.

$$s(AX) = \sqrt{\Delta x_1^2 + \ldots + \Delta x_n^2} + \varepsilon,$$

где ε — величина малая в сравнении с первым слагаемым и притом тем меньшая, чем меньше разности координат $\Delta x_1, \ldots, \Delta x_n$ [1].

Это и есть точное математическое определение римановой метрики и риманова пространства. Отличие римановой метрики, т. е. закона измерения расстояний, от эвклидовой состоит в том, что такой закон имеет место только вблизи каждой данной точки. Кроме того, координаты, в которых он так просто выражается, приходится брать разные для разных точек [2]. Отличие общей римановой метрики от эвклидовой мы еще уточним дальше.

То, что риманово пространство совпадает в бесконечно малом с эвклидовым, позволяет определить в нем основные геометрические величины, подобно тому как это делается во внутренней геометрии поверхностей посредством приближения бесконечно малого куска поверхности плоскостью (том 2, глава VII, § 4). Например, бесконечно малый объем выражается так же, как в эвклидовом пространстве. Объем конечной области получается суммированием бесконечно малых объемов, т. е. интегрированием дифференциала объема. Длина кривой определяется суммированием бесконечно малых расстояний между ее бесконечно близкими точками, т. е. интегрированием вдоль кривой дифференциала длины ds. Это и есть строгое аналитическое выражение того, что длина определяется откладыванием малого (бесконечно малого) масштаба вдоль кривой. Угол между кривыми, исходящими из одной точки, определяется точно так же, как в эвклидовом пространстве. Далее, в n-мерном римановом пространстве можно задавать поверхности разного числа измерений от 2 до $n-1$. При этом легко доказывается, что каждая такая поверхность представляет в свою очередь некоторое риманово пространство соответствующего числа измерений, совершенно подобно тому, как поверхность в обычном эвклидовом пространстве оказывается двумерным римановым пространством.

---

[1] Обычно точный смысл формулы (12) выражают следующим образом. Пусть из точки A исходит кривая, так что координаты ее точек X заданы как функции $x_1(t), x_2(t), \ldots, x_n(t)$ некоторой переменной t. Тогда дифференциал ds длины дуги этой кривой в точке A выражается формулой (12).

[2] Если бы можно было во всем пространстве ввести координаты так, что для любой пары близких точек имел бы место такой закон измерения расстояний, то пространство было бы эвклидовым.

Доказано также, что риманово пространство всегда может быть представлено поверхностью в эвклидовом пространстве достаточно большого числа измерений, именно: для всякого n-мерного риманова пространства найдется в $\frac{n(n+1)}{2}$-мерном эвклидовом пространстве n-мерная поверхность, которая с точки зрения ее внутренней геометрии не отличается от этого риманова пространства (по крайней мере в данной ее ограниченной части).

4. Для того чтобы действительно получить в римановой геометрии аналитические выражения разных геометрических величин, нужно прежде всего дать общее выражение для закона измерения длин в римановом пространстве, не связанное специально с особыми для каждой точки координатами. Ведь формула (12) имеет место для каждой точки A при специальном для этой точки выборе координат, так что при переходе от одной точки к другой нужно менять и сами координаты, а это, конечно, неудобно. Избавиться от этого, оказывается, легко, именно можно доказать следующее.

Пусть в какой-либо области риманова пространства введены какие угодно координаты $y_1, y_2, \ldots, y_n$. Тогда «бесконечно малое расстояние», или, как говорят, «элемент длины» ds от точки A с координатами $y_1, y_2, \ldots, y_n$ до точки X с координатами $y_1 + dy_1, y_2 + dy_2, \ldots, y_n + dy_n$ выражается формулой

$$ds = \sqrt{\sum_{i,k=1}^{n} g_{ik} dy_i dy_k}, \qquad \text{или} \qquad ds^2 = \sum g_{ik} dy_i dy_k, \tag{13}$$

где коэффициенты g_{ik} являются функциями координат $y_1, y_2, \ldots, y_n$ точки A.

Выражение, стоящее в последней формуле справа, называется квадратичной формой относительно дифференциалов координат $dy_1, \ldots, dy_n$[1]. В развернутом виде она пишется так:

$$\sum g_{ik} dy_i dy_k = g_{11} dy_1^2 + g_{12} dy_1 dy_2 + g_{21} dy_2 dy_1 + g_{22} dy_2^2 + \ldots$$

Так как $dy_1 dy_2 = dy_2 dy_1$, то второй и третий члены удобно считать одинаковыми: $g_{12} = g_{21}$ и вообще $g_{ik} = g_{ki}$; это возможно, поскольку важна только их сумма $(g_{ik} + g_{ki})\, dy_i dy_k$.

Квадратичная форма (13) положительна, так как, очевидно, $ds^2 > 0$, кроме того случая, когда все дифференциалы dy_i равны нулю.

Имеет место также обратное. Именно, если в n-мерном пространстве, где введены координаты $y_1, y_2, \ldots, y_n$, задать элемент длины формулой (13) с тем условием, что стоящая там квадратичная форма положительна (т. е. всегда больше нуля, кроме того случая, когда все

[1] Квадратичной формой от некоторых величин называют алгебраическое выражение, являющееся по отношению к этим величинам однородным многочленом 2-й степени.

$dy_i = 0$), то пространство будет римановым. Иными словами, вблизи каждой точки A можно будет ввести новые специальные координаты $x_1, x_2, \ldots, x_n$ так, что в новых координатах в этой точке элемент длины будет выражаться простой формулой (12)

$$ds^2 = dx_1^2 + dx_2^2 + \ldots + dx_n^2.$$

Таким образом, риманова метрика (т. е. определение длин, эвклидовское в бесконечно малом) может задаваться любой положительной квадратичной формой (13) с коэффициентами g_{ik}, являющимися функциями координат y_i. Таков общий метод задания римановой метрики.

Кривая в римановом пространстве задается тем, что все n координат точки меняются в зависимости от одного параметра t, изменяющегося в некотором промежутке

$$y_1 = y_1(t),\ y_2 = y_2(t), \ldots,\ y_n = y_n(t) \quad (a \leqslant t \leqslant b). \tag{14}$$

Длина кривой выражается интегралом

$$s = \int ds = \int \sqrt{\sum g_{ik} dy_i dy_k}\,.$$

В случае кривой, заданной уравнениями (14),

$$dy_1 = y_1' dt, \ldots,\ dy_n = y_n' dt,$$

поэтому

$$s = \int_a^b \sqrt{\sum g_{ik} y_i' y_k'}\, dt. \tag{15}$$

Так как g_{ik} являются известными функциями координат $y_1, \ldots, y_n$, а эти последние зависят известным образом от t по формулам (14), то в формуле (15) под знаком интеграла стоит вполне определенная для данной кривой функция от t. Стало быть, интеграл от нее имеет определенное значение, и тем самым кривая имеет определенную длину.

Длина кратчайшей из кривых, соединяющих две данные точки A, B, принимается за расстояние между этими точками. Сама эта кривая — геодезическая — играет роль аналога прямолинейного отрезка AB. Можно доказать, что в малой области любые две точки соединяются единственной кратчайшей линией. Сама задача нахождения геодезических (кратчайших) линий есть задача о придании минимума интегралу (15). Это есть задача вариационного исчисления, с которым читатель мог познакомиться в главе VIII (том 2). Стандартное применение методов вариационного исчисления позволяет вывести дифференциальное уравнение, определяющее геодезические

линии, и установить их общие свойства для любого риманова пространства.

Докажем высказанное ранее основное утверждение, что риманова метрика в любых координатах задается общей формулой (13).

Пусть в некоторой области риманова пространства введены какие-либо координаты $y_1, y_2, \ldots, y_n$. Возьмем в этой области произвольную точку A, и пусть $x_1, x_2, \ldots, x_n$ — те специальные для точки A координаты, в которых элемент длины в этой точке выражается формулой (12) или, что то же самое,

$$ds^2 = dx_1^2 + dx_2^2 + \ldots + dx_n^2. \tag{16}$$

Координаты x_i выражаются через координаты y_j $(i, j = 1, \ldots, n)$ какими-то формулами

$$\begin{aligned} x_1 &= f_1(y_1, y_2, \ldots, y_n), \\ x_2 &= f_2(y_1, y_2, \ldots, y_n), \\ &\cdots\cdots\cdots\cdots \\ x_n &= f_n(y_1, y_2, \ldots, y_n). \end{aligned}$$

Тогда

$$dx_1 = \frac{\partial f_1}{\partial y_1} dy_1 + \frac{\partial f_2}{\partial y_2} dy_2 + \ldots + \frac{\partial f_n}{\partial y_n} dy_n$$

и аналогично для $dx_2, \ldots, dx_n$. Подставим эти выражения в формулу (16). Тогда, возводя правые части в квадрат и объединяя члены с dy_1^2, $dy_1 dy_2$, dy_2^2 и т. д., получим выражение вида

$$ds^2 = g_{11} dy_1^2 + 2g_{12} dy_1 dy_2 + g_{22} dy_2^2 + \ldots + g_{nn} dy_n^2,$$

где коэффициенты $g_{11}, g_{12}, \ldots g_{nn}$ как-то выражаются через частные производные $\frac{\partial f_i}{\partial y_j}$ (вид этих выражений для нас несущественен). Но это и есть не что иное, как формула (13), написанная в развернутом виде, и тем самым наше утверждение доказано.

Докажем теперь, что, обратно, формула (13) определяет риманову метрику, т. е. что при специальном для каждой точки выборе координат x_i ее можно привести к простому виду (16). Пусть

$$ds^2 = \sum g_{ik} dy_i dy_k,$$

причем g_{ik} являются функциями координат $y_1, \ldots, y_n$, и стоящая справа квадратичная форма положительна. Возьмем какую-либо точку A и рассмотрим данную форму только в этой точке. Тогда коэффициенты g_{ik} оказываются данными числами, а переменными, от которых зависит форма, будут $dy_1, \ldots, dy_n$. Из алгебры известно, что всякую положительную квадратичную форму (с любыми численными коэффициентами) можно путем линейного преобразования переменных привести к сумме квадратов (см. главу XVI) [1], т. е. существует такое преобразование

$$\begin{aligned} dy_1 &= a_{11} dx_1 + \ldots + a_{1n} dx_n, \\ &\cdots\cdots\cdots\cdots \\ dy_n &= a_{n1} dx_1 + \ldots + a_{nn} dx_n, \end{aligned} \tag{17}$$

что после подстановки этих выражений в формулу (13) получим

$$ds^2 = dx_1^2 + \ldots + dx_n^2.$$

[1] Не существенно, что у нас переменные в форме — дифференциалы: мы можем рассматривать их просто как какие-то независимые переменные.

Если мы сделаем замену координат $y_1, \ldots, y_n$ на $x_1, \ldots, x_n$ по формулам

$$y_1 = a_{11}x_1 + \ldots + a_{1n}x_n,$$
$$\ldots\ldots\ldots\ldots$$
$$y_n = a_{n1}x_1 + \ldots + a_{nn}x_n,$$

то дифференциалы dy_j выразятся через дифференциалы dx_i как раз по формулам (17). Стало быть, такая замена координат решает поставленную задачу: в координатах $x_1, \ldots, x_n$ в выбранной нами точке квадрат дифференциала ds^2 выражается в простом «эвклидовском» виде (16). Тем самым доказано, что общая формула (13) действительно задает риманову метрику.

5. Эвклидово пространство есть простейший частный случай риманова пространства[1]. Важной задачей римановой геометрии было дать аналитическое выражение отличия общего риманова пространства от эвклидова, определить, так сказать, меру неэвклидовости риманова пространства. Этой мерой служит так называемая кривизна пространства.

Нужно прежде всего подчеркнуть, что понятие о кривизне пространства вовсе не связано с представлением о том, что пространство находится в каком-либо высшем объемлющем пространстве, где оно как-то искривлено. Кривизна определяется внутри самого данного пространства и выражает его отличие от эвклидова в смысле его внутренних геометрических свойств. Это нужно ясно понимать, чтобы не связывать понятие о кривизне пространства ни с чем посторонним. Когда говорят, что наше реальное пространство имеет кривизну, это значит лишь, что его геометрические свойства отличны от свойств эвклидова пространства. Но это вовсе не означает, что наше пространство лежит в каком-то высшем пространстве, где оно как-то искривлено. Такое представление не имеет никакого отношения к применению римановой геометрии к реальному пространству, а принадлежит области спекулятивных фантазий.

Понятие кривизны риманова пространства обобщает на n измерений понятие гауссовой кривизны поверхности. Как было сказано в § 4 главы VII (том 2), гауссова кривизна служит мерой отклонения внутренней геометрии поверхности от геометрии на плоскости и может рассматриваться с чисто внутренне-геометрической точки зрения. Она

---

[1] В эвклидовом пространстве элемент длины в прямоугольных координатах выражается формулой (16): $ds^2 = \sum dx_i^2$. Если перейти к другим координатам, то, согласно выводам п. 4, ds^2 выразится некоторой квадратичной формой (13). Стало быть, в *любых* координатах и в эвклидовом пространстве имеет место такая же общая формула (13) для элемента длины. Эвклидово пространство отличается, однако, от любого другого тем, что в нем можно ввести координаты (и это будут прямоугольные координаты) так, что формула (16) при одних и тех же координатах будет выполняться всюду, а не только вблизи той или иной точки, как это имеет место в общем римановом пространстве.

есть не что иное, как кривизна того двумерного риманова пространства, которое представляет собою данная поверхность.

Напомним для примера две формулы внутренней геометрии, в которые входит гауссова кривизна. Пусть на поверхности вблизи какой-либо точки O имеется малый треугольник, стороны которого — геодезические линии; пусть его углы будут α, β, γ, а площадь σ. Величина $\alpha+\beta+\gamma-\pi$ выражает отличие суммы его углов от суммы углов треугольника на плоскости.

Если треугольник стягивается к точке O, то отношение $\alpha+\beta+\gamma-\pi$ к его площади σ стремится к гауссовой кривизне K в точке O. Иными словами, для малого треугольника

$$\frac{\alpha+\beta+\gamma-\pi}{\sigma}=K+\varepsilon,$$

где $\varepsilon\to 0$, когда треугольник стягивается к точке O. Это как раз показывает, что гауссова кривизна K служит мерой отличия суммы углов треугольника на поверхности от суммы углов треугольника на плоскости.

Рассмотрим еще малую окружность на поверхности с центром в точке O (т. е. геометрическое место точек, равноудаленных от O в смысле расстояния по поверхности). Если r — радиус окружности, а l — ее длина, то

$$l=2\pi r-\frac{\pi}{3}Kr^3+\varepsilon,$$

где K — опять гауссова кривизна в точке O, а ε обозначает величину, малую по сравнению с r^3.

Здесь гауссова кривизна выступает как мера отклонения длины малой окружности от величины $2\pi r$, которой она равна в эвклидовой геометрии.

Аналогичную роль играет кривизна риманова пространства. Она может быть определена, например, следующим образом. В данном римановом пространстве проводится гладкая поверхность F, образованная геодезическими линиями, проходящими через данную точку O. Гауссова кривизна этой поверхности и принимается за кривизну пространства в точке O в направлении поверхности F. Вообще говоря, эта кривизна будет различной не только в разных точках O, но и для разных геодезических поверхностей F, проходящих через одну и ту же точку O. Кривизна пространства в данной точке не характеризуется, стало быть, одним числом. Еще Риман ввел общий закон, связывающий кривизны разных поверхностей F в одной и той же точке. Благодаря этим связям кривизна в точке вполне характеризуется некоторой системой чисел — так называемым *тензором кривизны*.

Однако мы не можем вдаваться здесь в объяснения по этому поводу, так как это потребовало бы привлечения большого математического аппарата. Важно усвоить только, что кривизна есть мера неэвклидовости риманова пространства; она определяется внутри него самого как мера отклонения его метрики от метрики эвклидова пространства. Она определяет, например, отличие суммы углов треугольника от π и отличие длины окружности от $2\pi r$. В разных точках она имеет, вообще говоря, различные значения, но и в одной точке она задается не одним числом, а некоторой системой чисел.

Риманово пространство может не быть однородным по своим свойствам, и тогда в нем невозможно свободное движение фигур без изменения расстояний между их точками. Встает вопрос о тех римановых пространствах, в которых свободное движение фигур возможно и притом с тем же числом степеней свободы, что и в эвклидовом пространстве. Это будут наиболее однородные римановы пространства.

Оказывается, что эвклидово пространство есть однородное пространство без кривизны (пространство нулевой кривизны). Другой тип однородного пространства представляет пространство Лобачевского, так что геометрия Лобачевского, так же как геометрия Эвклида, оказывается частным случаем общей римановой геометрии.

Вообще римановы пространства, в которых возможно свободное движение фигур, — это пространства постоянной кривизны: в них кривизна имеет одно и то же значение во всех точках и для всех геодезических поверхностей. (Вместо «тензора кривизны», меняющегося от точки к точке, она задается на этот раз одним числом, общим для всех точек.) Пространство нулевой кривизны будет эвклидовым, пространство отрицательной кривизны — пространством Лобачевского; пространство положительной кривизны имеет ту же геометрию, что и n-мерная сфера в $(n+1)$-мерном эвклидовом пространстве.

6. Приложения римановой геометрии не заставили себя долго ждать. Сам Риман, как мы уже говорили, применил ее формальный аппарат к решению задач теплопроводности, но это было пока применение лишь ее формул, а не идеи абстрактного пространства с эвклидовским измерением расстояний в бесконечно малых областях. Такое приложение было дано к цветному пространству, где расстояние между близкими цветами стали выражать, пользуясь римановой метрикой; пространство цветов стали трактовать как особое трехмерное риманово пространство.

Появилось также важное приложение римановой геометрии в механике. Чтобы понять его сущность, рассмотрим сначала движение точки по поверхности. Представим себе материальную точку, например дробинку, которая может свободно двигаться по некоторой гладкой поверхности, но не покидает эту поверхность. Точка как бы движется в самой поверхности. На поверхности можно ввести какие-либо коор-

динаты x_1, x_2; тогда движение точки вполне определяется зависимостью этих ее координат от времени, а скорость — скоростями изменения координат, т. е. их производными по времени $\dot{x}_1$, $\dot{x}_2$. Получается, что точка как бы движется в двумерном пространстве, но это пространство не эвклидово, а имеет свою геометрию — внутреннюю геометрию поверхности. Законы движения можно преобразовать так, чтобы в них входили только координаты x_1, x_2 точки на поверхности, их первые и вторые производные.

Если на точку действует сила, то ее составляющая, перпендикулярная поверхности, погашается сопротивлением — реакцией поверхности и остается лишь составляющая, касательная к поверхности; она действует уже только вдоль поверхности. Таким образом, и силы, действующие на точку, можно считать действующими в самой поверхности. Внутренняя геометрия поверхности есть частный случай римановой геометрии. Поэтому движение точки по поверхности есть движение в двумерном римановом пространстве. Законы ее движения имеют тот же характер, что обычные законы движения, с той лишь разницей, что в них учитывается внутренняя геометрия поверхности. Это становится особенно ясным из следующего факта, упоминавшегося в § 4 главы VII (том 2): точка, движущаяся на поверхности по инерции и без трения, движется по геодезической линии с постоянной скоростью. Так как геодезические линии играют на поверхности роль прямых, то указанный факт аналогичен закону инерции; это — тот же закон инерции, но для движения на поверхности или, абстрактно, для движения в двумерном римановом пространстве.

Конечно, в этом абстрактном представлении пока не видно никаких преимуществ, поскольку речь идет все же о движении по обычной поверхности.

Выгода этой абстрактной точки зрения обнаружится сразу, когда мы перейдем к механическим системам, расположение которых задается более чем двумя величинами. Тогда изображение их движения движением точки на поверхности станет невозможным. С таким обстоятельством мы уже встречались в § 7, когда говорили о том, как графические методы перерастают в абстрактное представление о многомерном пространстве.

Итак, пусть имеется механическая система, конфигурация которой, т. е. расположение ее частей, задается n величинами x_1, x_2, ..., x_n. Если речь идет о системе из нескольких материальных точек, то расположение их определяется заданием всех их координат, по три на каждую точку. Другим примером может служить волчок (колесо, вращающееся на оси, которая сама может вращаться вокруг неподвижной точки). Поворот волчка вокруг оси задается углом поворота, а наклон оси — двумя углами, которые она образует с двумя данными

направлениями. Получается всего три величины, определяющие положение такого волчка (рис. 32).

Каждую конфигурацию — расположение частей системы — можно мыслить как «точку» пространства всех возможных ее конфигураций. Это будет так называемое конфигурационное пространство системы[1]. Число его измерений равно числу величин $x_1, x_2, \ldots, x_n$, определяющих конфигурацию. Эти величины служат координатами «точки» в конфигурационном пространстве. Для системы, скажем, из трех материальных точек получим по три координаты у трех точек — всего девять координат. Для случая волчка имеем три координаты — три угла, так что конфигурационное пространство волчка трехмерно.

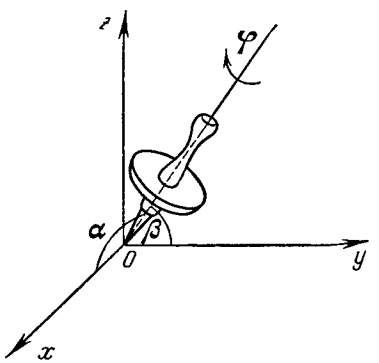


Рис. 32.

Движение системы представляется движением точки в конфигурационном пространстве. Скорость этого движения определяется скоростями изменения координат $x_1, x_2, \ldots, x_n$.

О таких пространствах в связи с их топологической структурой еще будет идти речь в главе XVIII. Здесь же мы хотим подчеркнуть, что в конфигурационном пространстве можно ввести специальный закон измерения расстояний, тесно связанный с механическими свойствами системы. Именно, если кинетическая энергия системы выражается формулой

$$T = \frac{1}{2} \sum_{i,k=1}^{n} a_{ik} \dot{x}_i \dot{x}_k,$$

где $\dot{x}_i$ — скорость изменения соответствующей координаты, то квадрат бесконечно малого расстояния задается формулой

$$ds^2 = \sum_{i,k=1}^{n} a_{ik} dx_i dx_k.$$

Таким образом, конфигурационное пространство становится римановым пространством. При этом движение системы не только представляется движением точки в конфигурационном пространстве, но и самые уравнения, описывающие движение системы, совпадают с уравнениями движения этой точки; словом, механика системы изображается как механика точки в конфигурационном пространстве. В частности, движение системы по инерции, т. е. без воздействия сил

[1] Не следует смешивать его с «фазовым пространством», упомянутым в § 8, п. 2. В фазовом пространстве точка определяет не только расположение, но и скорости движения точек системы в каждый момент.

(подобно свободному вращению волчка), представляется равномерным движением точки по геодезической линии в этом пространстве.

Это изображение оказывается в ряде случаев полезным, и им, как и некоторыми его обобщениями и видоизменениями, пользуются в теоретической механике.

Таким образом, риманова геометрия находит себе применение как метод абстрактно-геометрического описания физических явлений. Это описание вовсе не произвольно и не является игрой математического ума, оно отражает реальные закономерности рассматриваемых явлений, но отражает их в абстрактной форме. Такова, однако, природа всякого математического описания физических явлений. Такова же природа всякого применения абстрактной геометрии. Разница состоит лишь в том, что здесь применяются более сильные, более тонкие абстракции, но сущность остается той же самой.

Наиболее блестящее приложение риманова геометрия нашла в теории относительности. Но об этом мы будем говорить в следующем параграфе, потому что здесь речь идет о важном и трудном вопросе отношения абстрактной геометрии к свойствам реального пространства.

7. За последние 30 лет геометрия различных неэвклидовых пространств подверглась значительному развитию и обобщениям в разных направлениях. Возникли новые теории, в которые риманова геометрия была включена как частный случай. Первой из них была так называемая финслерова геометрия, идея которой восходит еще к Риману[1]; затем — общая теория пространств крупнейшего французского геометра Э. Картана, соединившего римапову геометрию с Эрлангенской программой Клейна, и другие теории. Не имея возможности говорить об этих новейших направлениях геометрии, заметим только, что они разрабатываются в основном посредством специально приспособленного для них аналитического аппарата. В развитии этих новых направлений участвует группа советских геометров; тут можно было бы назвать новую «полиметрическую» геометрию, созданную П. К. Рашевским, исследования В. В. Вагнера, простирающиеся от наиболее общих проблем теории кривых пространств до приложений неэвклидовой геометрии в механике, и др.

§ 10. АБСТРАКТНАЯ ГЕОМЕТРИЯ И РЕАЛЬНОЕ ПРОСТРАНСТВО

1. Следуя в предыдущем изложении развитию геометрических идей начиная от Лобачевского, мы углубились в абстрактные пространства и довольно далеко ушли от первоначального предмета геометрии — того реального пространства, в котором протекают все явления. Теперь мы обратимся к этому пространству в обычном смысле, и нашей задачей

[1] Финслер — немецкий математик, начавший в 1916 г. детальную разработку упомянутой геометрии.

будет выяснить, что дало развитие абстрактной геометрии для познания его свойств.

Мы знаем, что геометрия возникла из опыта, из опытного исследования пространственных форм и отношений тел: из измерений земельных участков, объемов сосудов и т. п. Таким образом, по происхождению она есть такая же физическая теория, как, скажем, механика. Аксиомы эвклидовой геометрии — это четко сформулированные выводы из длительного опыта; они выражают законы природы, и их можно называть законами геометрии так же, как основные законы механики часто теперь называют аксиомами механики [1]. Но нельзя утверждать, что эти законы являются абсолютно точными и никогда не потребуют уточнения и обобщения в связи с новыми опытными данными; реальные свойства пространства могут в той или иной мере отличаться от того, что дает эвклидова геометрия.

Эти рассуждения мы уже приводили, и теперь они представляются, пожалуй, совсем очевидными. Но не так было сто лет назад, когда идеи Лобачевского еще не завоевали общего признания. До Лобачевского и Гаусса никому и в голову не приходило, что эвклидова геометрия может оказаться не совсем точной, что реальные свойства пространства могут быть несколько иными. Лобачевский развивал свою геометрию как теорию возможных свойств реального пространства. Позже Риман и некоторые другие ученые также ставили вопрос о возможных свойствах пространства, о возможных законах измерения длин, которые могли бы обнаружиться при уточнении измерений. Вообще абстрактная геометрия в некоторых своих частях могла и может рассматриваться как теория возможных свойств пространства. Все это оставалось, однако, в области гипотез, пока в 1915 г. Эйнштейн в своей общей теории относительности не подтвердил идей Лобачевского и Римана. Эта теория утверждает, что геометрия реального пространства действительно несколько отлична от эвклидовой, и это обнаруживается именно в астрономических масштабах, как того ожидал Лобачевский.

В том, что было только-что сказано о пространстве, скрыты по крайней мере три трудности. К их выяснению и сводится в конечном счете проблема отношения абстрактной геометрии к физической геометрии, т. е. к свойствам реального пространства.

Первая трудность состоит в том, чтобы по возможности представить себе, как и в каком смысле свойства реального пространства могут отли-

[1] Абстрактное понимание аксиом, отвлекающее аксиомы от их первоначального содержания, возникло всего лет пятьдесят назад, и оно ничего не меняет в том факте, что аксиомы эвклидовой геометрии выражают законы природы. Говоря об аксиомах, а не о законах геометрии или механики, хотят выдвинуть на первый план логическое дедуктивное построение этих наук, но от этого эти науки не теряют своего опытного основания. Какое-либо положение теории называется аксиомой, если оно принято за основу при дедуктивном построении теории, когда другие положения теории (теоремы) выводятся из основных (аксиом) путем логических рассуждений.

чаться от того, что дает эвклидова геометрия. Мы слишком привыкли к ней, чтобы легко вообразить себе что-нибудь другое, и разъяснения тут, конечно, необходимы.

Вторая трудность заключена в самом выражении «свойства реального пространства». Пространство само по себе мыслят как пустое и однородное. Казалось бы, в самом понятии пространства уже заключено представление о его однородности. И как пустое пространство, т. е. «пустота», может иметь какие-то свойства? Мы говорим «свойства пространства», не задумываясь над этими вопросами, но стоит в них вдуматься, как указанная трудность станет достаточно ощутимой.

Третья трудность заключена в понятии об истинности той или иной геометрии. Казалось бы, дело очень просто: та геометрия истинна, которая соответствует действительности. Это, конечно, так. Но, с другой стороны, мы видели, например, что геометрию внутри круга можно понимать как геометрию Лобачевского, поскольку всякий геометрический факт внутри круга можно изложить как теорему геометрии Лобачевского. Следовательно, оказывается, что одни и те же геометрические факты можно излагать и как теоремы геометрии Эвклида и как теоремы геометрии Лобачевского. Значит, обе геометрии соответствуют действительности. Так, какая же из них истинна и в каком смысле, и почему мы все-таки считаем, что в круге на самом деле выполняется эвклидова геометрия, а геометрия Лобачевского в нем лишь изображается, интерпретируется?

Ясно, что в этих вопросах имеется известная трудность, о которую в свое время споткнулись и некоторые крупные математики.

Разъяснения следует начать со второй из названных трудностей, потому что из понимания того, что такое «свойства пространства», последует решение и других затруднений.

2. Предмет геометрии — «свойства пространства» — составляют свойства реальных тел, их материальные отношения и формы. В реальном пространстве «место», «точка», «направление» и т. п. определяется материальными телами. «Здесь» и «там», «туда» и «сюда» и т. п. имеют смысл лишь в связи, в отношении к тем или иным материальным предметам. «Здесь» может означать «на земле», «в этой комнате» или что-нибудь в том же роде; словом, «здесь» — это всегда означает место, определенное теми или иными материальными признаками. Точно так же, например, прямая линия не существует сама по себе, а лишь как натянутая нить, или край линейки, или луч света. Прямая же вообще, «прямая как таковая», есть абстракция, отражающая общие свойства этих материальных прямых, точно так же, как, скажем, «дом вообще» есть абстракция, отражающая общие свойства домов; «дом вообще» не существует вне и независимо от отдельных реальных домов.

Этот объективный характер свойств пространства выражен известным положением диалектического материализма: *пространство есть форма существования материи*. Форма предмета определяется связями и отно-

шениями его частей. Структура пространства есть общая закономерность ряда отношений материальных тел и явлений. Это — пространственные отношения, пространственный порядок предметов, их взаимное расположение, расстояния и т. п. Но как всякая форма неотделима от содержания иначе, как в абстракции и в известных рамках, так и пространство неотделимо от материи. Представление о пространстве «самом по себе», пространстве без материи, есть абстракция, которой нельзя злоупотреблять. Реальные пространственные отношения и формы: «здесь», «между», «внутри», «прямая», «шар» и т. д. и т. п. — это всегда отношения и формы материальных тел. Геометрия же рассматривает их отвлеченно. Это отвлечение от конкретности необходимо, ибо иначе нет возможности познать общее в разнообразных конкретных отношениях предметов. Но нельзя это отвлечение абсолютизировать, подменять отвлеченными понятиями саму объективную реальность.

В абсолютно пустом пространстве, вовсе лишенном следов материи, ничем не различались бы ни места, ни направления, тут, следовательно, нет ни места, ни направлений, так что абсолютно пустое пространство превращается в ничто. Даже в абстрактном представлении пустого пространства молчаливо подразумевают, что в нем различимы разные места и направления. Другими словами, в отвлеченном представлении о пространстве удерживают свойства различимости места, направления, расстояния, которые в реальном пространстве существуют именно потому, что это пространство неразрывно связано с материальными телами.

Итак, пространство есть форма существования материи; «свойства пространства» — это, следовательно, свойства материи, свойства известных отношений материальных тел, их взаимного расположения, размеров и т. п.

Далее, для того чтобы теоремы геометрии имели физический смысл, нужно знать, что в них следует понимать под «прямой», «расстоянием» и другими геометрическими понятиями. В § 4 мы видели, что одна и та же геометрическая теория допускает разные толкования.

Следовательно, сравнивая геометрию с опытом, нужно по возможности точно определять физический смысл геометрических понятий, поскольку геометрия описывает свойства реального пространства лишь при том условии, что ее понятиям приписывается соответствующий физический смысл. Без этого физического смысла теоремы геометрии имеют абстрактно математический, формальный характер. В этом и состоит решение указанного выше второго затруднения. Это затруднение возникает потому, что, вместо реального пространства, неразрывно связанного с материей, хотят мыслить «чистое» пространство, пространство «как таковое», которое есть, однако, не более как абстракция.

3. Теперь легко будет понять, как решаются два других затруднения.

Как, во-первых, можно представить себе, что реальное пространство по своим свойствам отлично от эвклидова? Пусть мы хотим проверить

какое-либо утверждение эвклидовой геометрии, например, то, что сумма углов треугольника равна 180°, или то, что длина окружности равна $2\pi R$. Для проверки первого нужно определить, какие физически определенные треугольники будут рассматриваться и как будут определяться их углы. Пусть сторона треугольника — это луч света в пустоте. В таком случае нет ничего невероятного в том, что очень точные опыты покажут отличие суммы углов треугольника от 180°. Точно так же можно мыслить, что измерение радиуса и окружности одним и тем же масштабом приведет к результатам, не удовлетворяющим в точности соотношению $l=2\pi R$. Кстати, так это и есть на земной поверхности, где длина окружности не пропорциональна радиусу, а растет медленнее и достигает максимума, когда радиус делается равным половине меридиана. На это можно возразить, что ведь на земной поверхности роль прямых играют дуги больших кругов, так что радиус понимается здесь в другом смысле, и потому наш результат не противоречит эвклидовой геометрии. Однако, согласно теории относительности, вблизи тел большой массы отношение длины окружности к «настоящему» радиусу все-таки несколько отличается от 2π, а именно верна следующая приближенная формула для отношения длины экватора однородного шарообразного тела к его радиусу:

$$\frac{\text{длина окружности}}{\text{радиус}} = 2\pi\left(1 - \frac{kM}{3Rc^2}\right),$$

где M — масса тела (в т), R — радиус тела (в км), $c=300\,000$ км/сек — скорость света, k — постоянная тяготения, равная при указанном выборе единиц измерения $66{,}6 \cdot 10^{-18}$.

Таким образом, мы видим, что отношение длины окружности к радиусу не равно 2π, а несколько меньше. Как показывает вычисление, на поверхности Солнца это отношение отличается от 2π приблизительно на 0,000 004, а на поверхности спутника Сириуса, средняя плотность которого в 50 000 раз больше плотности воды, отклонение достигает уже 0,00014.

Можно, конечно, возразить, что все это тем не менее нельзя себе представить, что в наглядном представлении пространство всегда будет эвклидовым. Такое возражение не должно нас смущать, во-первых, потому, что задача науки состоит не в том, чтобы дать наглядное представление явлений, но в том, чтобы достигнуть их понимания. Наглядное представление ограничено и вращается в кругу привычных образов, доставляемых нашими органами чувств. Поэтому мы не в состоянии наглядно представить ни ультрафиолетовых лучей, ни распространения радиоволн, ни движения электрона в атоме, ни многого другого, иначе как подставляя на место этих явлений какие-нибудь модели. Но это вовсе не значит, что эти явления нам непонятны. Напротив, успехи радиотехники, например, ясно показывают, что мы вполне овладели радиовол-

нами и, следовательно, понимаем их достаточно хорошо. Во-вторых, решение вопроса о том, что можно себе представить, а чего нельзя, зависит от привычки и тренировки воображения. Можно ли представить себе антиподов, которые по отношению к нам ходят вниз головой? Теперь мы можем себе это представить, а в свое время «невообразимость» антиподов служила аргументом против шарообразности Земли.

4. Теперь обратимся к последней трудности, т. е. к вопросу о том, какую геометрию следует считать истинной. При самой постановке этого вопроса было указано, что геометрические факты внутри круга можно излагать как теоремы и геометрии Эвклида и геометрии Лобачевского. Следовательно, обе эти геометрии отвечают действительности, т. е. обе они истинны. И в этом нет, в конце концов, ничего удивительного. Одно и то же явление всегда можно описывать разными способами; одну и ту же величину можно измерять разными единицами; одну и ту же кривую можно задавать разными уравнениями в зависимости от выбора системы координат. Совершенно так же данную выделенную совокупность геометрических соотношений (как в рассматриваемом примере соотношения внутри круга) можно описывать разными способами. Но мы ставим вопрос не о какой-то изолированной совокупности геометрических фактов, а о пространственных отношениях во всей их общности. Пространство есть универсальная форма существования материи, и, стало быть, при постановке вопроса о свойствах пространства никакая область фактов не может быть искусственно выделена.

При такой постановке вопроса геометрические величины, геометрические факты нельзя рассматривать изолированно от других явлений, с которыми они необходимо связаны. Например, длина отрезка определяется путем откладывания твердого масштаба, так что измерение длин необходимо связано с движением твердых тел, т. е. с механикой. Геометрия неотделима от механики. Между тем измерение длин внутри круга при интерпретации геометрии Лобачевского производится совсем иначе, как это было указано в § 4; хорда получает при этом бесконечную длину. Ясно, что так определенное измерение не соответствует первоначальному представлению об измерении, возникшему именно на основе механического перемещения сравниваемых реальных тел. Вообще фигуры, равные в смысле эвклидовой геометрии, это, по самому происхождению эвклидовой геометрии, те фигуры, которые накладываются друг на друга путем механического движения. В интерпретации геометрии Лобачевского равенство определяется иначе, роль движений там играют другие преобразования. Следовательно, если геометрические факты внутри круга брать в их необходимой связи с механикой, то мы должны признать, что на самом деле внутри круга имеет место (с большой точностью) именно эвклидова геометрия.

Эвклидова геометрия — это та, в которой роль движения играет обычное механическое движение твердых тел. Именно поэтому люди

открыли сперва эвклидову, а не какую-либо другую геометрию. Но развитие физики привело теперь к выводу, что законы ньютоновской механики и вместе с ними законы эвклидовой геометрии являются лишь приближением к более точным и общим законам. В этом изменении законов геометрии, в переходе от эвклидовой к римановой геометрии, осуществляемом в теории относительности, играла роль не только механика: не меньшее, если не большее, значение имела здесь теория электромагнитных явлений и оптика. Геометрия, как наука о свойствах пространства, связана с физикой, зависит от нее и может быть отделена от нее лишь в абстракции, в известных рамках.

Зависимость геометрии от физики, зависимость свойств пространства от материи была явно указана еще Лобачевским, который предвидел возможность изменения законов геометрии при переходе к новой области физических явлений. В противоположность этой материалистической точке зрения, известный математик Пуанкаре еще сравнительно недавно утверждал, будто выбор той или иной геометрии диктуется лишь соображениями простоты или «экономии мышления» в духе известного идеалиста Маха. На этом основании Пуанкаре предсказывал даже, что наука легче откажется от закона прямолинейного распространения света, чем от эвклидовой геометрии, поскольку она «самая простая». Однако Пуанкаре не дожил всего трех лет до окончательного построения общей теории относительности, в которой поступают как раз наоборот: отказываются от эвклидовой геометрии, но сохраняют закон распространения света, правда, в обобщенной форме: свет распространяется по геодезическим линиям. И это был не первый случай крушения махистских предсказаний. На пороге нашего столетия махисты и другие идеалисты близких толков утверждали, будто атомы являются лишь мысленными символами или чем-нибудь в этом роде, но никак не объективной реальностью. Между тем не прошло и нескольких лет, как реальность атомов была доказана самыми убедительными опытами. Несмотря на все это, идеалистическое понимание геометрии, как и других наук, еще сохраняется среди тех буржуазных ученых, для которых буржуазная идеология оказывается сильнее объективной истины.

Итак, мы приходим к следующим выводам. Одну и ту же выделенную совокупность фактов можно, вообще говоря, описывать по-разному, и все такие описания истинны, раз они отражают действительность. Однако геометрические факты во всей их общности никак нельзя рассматривать в отрыве от других явлений. Только так можно устанавливать свойства пространства, поскольку оно является универсальной формой существования материи. А беря геометрию в связи с физикой, мы необходимо должны приспосабливать их друг к другу, и тогда выявляется существенное различие между разными «геометриями», которые в отрыве от физики могут казаться отличающимися лишь большей или меньшей простотой. Эвклидова геометрия появилась не потому, что она проще других, а по-

тому, что она соответствует механике. Теперь же именно в связи с развитием физики в теории относительности переходят к более сложной геометрии — римановой.

Словом, в отношении к свойствам реального пространства истинна та геометрия, которая достаточно точно отражает свойства пространственных отношений во всей их общности и которая, стало быть, отвечает не только чисто геометрическим фактам, но также механике и вообще всей физике.

5. То немногое, что было сказано выше о теории относительности, не касается, однако, главного в ее содержании. В понимании проблемы пространства она пошла существенно дальше, чем об этом думали Лобачевский и Риман.

Наиболее существенное и основное положение теории относительности состоит в том, что пространство неотделимо полностью от времени, и что они вместе образуют единую форму существования материи: четырехмерное многообразие пространства-времени. События в мире характеризуются местом и временем, а следовательно, четырьмя координатами: тремя пространственными и четвертой временно́й — временем события. События образуют в этом смысле четырехмерную совокупность. Эту-то четырехмерную совокупность с точки зрения ее структуры, в отвлечении от свойств отдельных явлений и рассматривает прежде всего теория относительности. Она не является в своей основе и сущности ни теорией быстрых движений, ни космологией, ни новой теорией пространства или времени, — это именно теория пространства-времени как единой формы существования материи.

Конечно, и в ньютоновской механике можно объединить пространство и время в одно четырехмерное многообразие. Как мы уже имели случай упомянуть, сама идея многомерного пространства зародилась впервые у Лагранжа именно так, что, рассматривая движение материальной точки, он присоединял к пространственным координатам x, y, z еще временную t. Движение точки изображается тогда линией в четырехмерном пространстве с координатами x, y, z, t; при движении точки меняются все четыре координаты: положение (x, y, z) и время t. Однако тут объединение пространства и времени носит чисто формальный характер. Никакой внутренней необходимой связи между пространством и временем здесь не устанавливается. Конечно, в законе движения каждого данного тела есть своя зависимость пространственного положения от времени. Но это касается только каждого данного движения, никакой всеобщей внутренней связи между пространством и временем ни в механике, ни вообще в физике не было установлено до теории относительности. Всегда однозначно различались пространственные отношения, пространственный порядок вещей и явлений от их отношений и порядка во времени. Временная последовательность событий, продолжительность промежутков времени считались абсолютными, определенными безотносительно к чему

бы то ни было. Короче, в физике безраздельно господствовало понятие абсолютного времени.

Величайшим открытием Эйнштейна, составившим не только краеугольный камень теории относительности, но и поворотный пункт в общем физическом и философском понимании проблемы пространства и времени, было открытие того, что абсолютного времени в действительности нет. Вскоре после того, как Эйнштейн построил в 1905 г. свою теорию[1], Минковский показал, что суть ее состоит не столько в отказе от абсолютного времени, сколько в установлении взаимосвязи пространства и времени, в силу которой имеется единая абсолютная форма, существования материи: пространство-время. Отделение же пространства — пространственных координат — от времени, от временно́й координаты t является в известной мере относительным, зависящим от той материальной системы, — «системы отсчета», в отношении которой определяется пространственный и временно́й порядок явлений. События, одновременные по отношению к одной системе, могут не быть таковыми в отношении к другой системе.

В определении порядка явлений, конечно, нет и не может быть полной зависимости от системы отсчета. Порядок событий, связанных прямым взаимодействием, само собой разумеется, остается в отношении всех систем одним и тем же, так что действие всегда предшествует его результату. Но для событий, не связанных взаимодействием, порядок во времени оказывается относительным. Поскольку пространственный порядок (в его чистом виде) относится к одновременным событиям, а одновременность относительна, выделение чисто пространственных отношений из общей совокупности пространственно-временных отношений оказывается относительным, зависящим от системы отсчета. Пространство в абстрактном смысле оказывается как бы «сечением» четырехмерного многообразия пространства-времени, проведенным через одновременные (в отношении данной системы) события.

В нашу задачу не может входить изложение основ теории относительности, мы постараемся лишь в самых кратких словах охарактеризовать ее основы в том виде, как их представляется наиболее естественным понимать в связи с идеями абстрактной геометрии. Это понимание, кстати сказать, существенно отлично от того, которое исходит от самого Эйнштейна.

Мир, вселенную можно рассматривать как множество разнообразных событий. Под событием понимается при этом не какое-либо явление, простирающееся в пространстве и длящееся во времени, но как бы мгновенно-точечное явление, наподобие мгновенной вспышки точечной лампы. Говоря языком геометрии, события — это точки четырехмерного многообразия вселенной.

[1] Теория, построенная Эйнштейном в 1905 г., называется специальной теорией относительности в противоположность «общей», построенной Эйнштейном в 1915 г.

Пространство-время есть форма существования материи, форма этого мирового многообразия. Структура пространства-времени, его «геометрия», есть не что иное как некоторая общая стуктура мира, т. е., согласно нашему рассмотрению, «геометрия» множества событий. Структура эта определяется некоторыми всеобщими материальными взаимосвязями и отношениями событий.

Во-первых, как мы уже разъясняли выше для пространственных отношений, это должны быть именно материальные отношения и взаимосвязи. То же верно для отношений явлений во времени. Пространственные и временны́е отношения *сами по себе*, в чистом виде, суть лишь абстракции.

Во-вторых, отношения событий, определяющие структуру пространства-времени, должны иметь всеобщий характер согласно универсальному характеру пространства-времени.

Такое всеобщее материальное отношение событий представляет их причинно-следственная связь. Всякое событие так или иначе, прямо или косвенно воздействует на некоторые другие события и в свою очередь испытывает воздействие других событий. Это отношение воздействия одних событий на другие и определяет структуру пространства-времени.

Таким образом, теория относительности позволяет дать следующее определение. Пространство-время есть множество всех событий, отвлеченное от их конкретных свойств и отношений, кроме общего отношения воздействия одних событий на другие. Само это отношение так же следует понимать здесь в общем смысле, в отвлечении от его разнообразных конкретных форм.

В специальной теории относительности пространство-время считается максимально однородным. Это означает, что многообразие событий допускает преобразования, которые не нарушают отношения воздействия между событиями, и группа таких преобразований, в известном смысле, максимально возможная. Пусть, например, имеются две пары событий A, B и A', B', и как A не воздействует на B, так и B не воздействует на A, и аналогичное верно для A' и B'. Тогда существует такое соответствие между событиями, при котором A соответствует A', а B соответствует B', и при этом для любых пар событий отношение воздействия (или невоздействия) не нарушается, т. е. если X воздействует на Y, то соответствующее событие X' также воздействует на Y', и если X не воздействует на Y, то и X' не воздействует на Y'.

В соответствии с этим оказывается, что с точки зрения специальной теории относительности пространство-время есть своего рода четырехмерное пространство, геометрия которого определяется некоторой группой преобразований. Эти преобразования и есть не что иное как знаменитые преобразования Лоренца. Законы геометрии и физики не изменяются при этих преобразованиях. Такой взгляд соответствует взгляду на геомет-

рию, выдвинутому Клейном в его Эрлангенской программе, о которой говорилось в § 6.

6. Общая теория относительности идет дальше, она отказывается от представления об однородности пространства-времени. Она считает, что пространство-время однородно лишь с известным приближением в достаточно малых областях, но в целом неоднородно. Неоднородность пространства-времени определяется согласно теории Эйнштейна распределением и движением материи. В свою очередь структура пространства-времени определяет законы движения тел, и это обнаруживается в явлении всемирного тяготения. Общая теория относительности оказывается, собственно, теорией тяготения, объясняющей тяготение связью стуктуры пространства-времени с движением материи.

Представление о пространстве-времени, однородном лишь с известным приближением в малых областях, аналогично представлению о римановом пространстве, которое является эвклидовым лишь «в бесконечно малом». Математически пространство-время в общей теории относительности и трактуется как своего рода риманово пространство, правда, в существенно измененном смысле.

Именно, в четырехмерном римановом пространстве вблизи каждой точки можно ввести координаты так, что квадрат линейного элемента выражается формулой $ds^2 = dx_1^2 + dx_2^2 + dx_3^2 + dx_4^2$.

В пространстве-времени можно в окрестности каждого события ввести координаты x, y, z, t так, что линейный элемент представляется формулой $ds^2 = dx^2 + dy^2 + dz^2 - c^2dt^2$, где c — скорость света, которую при соответствующем выборе единиц измерения можно считать равной единице. Здесь x, y, z — пространственные координаты, а t — время. Минус при dt^2 дает формальное выражение существенного, коренного отличия временно́й координаты от пространственных, времени — от пространства.

В теории тяготения важнейшую роль играет понятие о кривизне пространства-времени. Основные уравнения этой теории, данные Эйнштейном, как раз связывают величины, характеризирующие кривизну пространства-времени, с величинами, характеризующими распределение и движение материи. Эти уравнения есть вместе с тем уравнения поля тяготения и из них же, как показали Эйнштейн с сотрудниками и В. А. Фок, выводятся законы движения тел в поле тяготения.

Структура пространства-времени по общей теории относительности оказывается сложной, и пространство нельзя отделить от времени даже в такой степени, как это допускает специальная теория относительности. Однако с известным приближением и при известных предположениях это можно сделать. Пространство оказывается эвклидовым с достаточной точностью в областях, малых в сравнении с космическими масштабами, но в больших областях обнаруживается отклонение от эвклидовой геометрии. Это отклонение зависит от распределения и движения масс

материи и достигает заметных, хотя все же очень малых, величин вблизи звезд большой массы, а также в целом в космических масштабах. В ряде гипотез о структуре вселенной в целом приближенно предполагают массы распределенными в среднем равномерно. В одной из таких гипотетических теорий, предложенных советским физиком и математиком Фридманом, геометрия пространства в целом совпадает с геометрией Лобачевского.

В теории относительности абстрактная геометрия находит свое применение не только как математический аппарат; самые идеи абстрактного пространства дают средство наиболее глубокой формулировки основ этой теории. Возможности, намеченные абстрактной геометрией, открываются в действительности, и теоретическое мышление празднует здесь свой блестящий триумф. Абстрактная геометрия, сама выросшая из опытного изучения пространственных отношений и форм тел, противостоит теперь изучению реального пространства как готовый математический метод. Таков общий путь науки: от непосредственно данного в опыте она поднимается к теоретическим обобщениям и абстракциям, которые опять обращаются к опыту, как инструмент более глубокого познания сущности явлений, давая объяснение известных и предсказание новых явлений, направляя практическую деятельность людей и находя в этом свое оправдание и источник дальнейшего развития.

ЛИТЕРАТУРА[1]

Александров А. Д. Геометрия. БСЭ, т. 10, стр. 533—550.

Делоне Б. Н. Краткое изложение доказательства непротиворечивости планиметрии Лобачевского. Изд. 2-е, Гостехиздат, 1956.

Широков П. А. и Каган В. Ф. Строение неэвклидовой геометрии, Гостехиздат, 1950.

Книга содержит доступное и сравнительно полное изложение геометрии Лобачевского.

Каган В. Ф. Геометрические идеи Римана и их современное развитие. ГТТИ, 1933.

Блохинцев Д. И. и Дрябкина С. И. Теория относительности А. Эйнштейна, Гостехиздат, 1940.

Фок В. А. Современная теория пространства и времени. «Природа», № 12, 1953.

Статья в возможно доступной форме освещает основы современных представлений о геометрии реального пространства.

Фок В. А. Теория пространства, времени и тяготения, Гостехиздат, 1955.

Систематические курсы

Ефимов Н. В. Высшая геометрия. Изд. 3-е, Гостехиздат, 1953.

Рашевский П. К. Риманова геометрия и тензорный анализ, Гостехиздат, 1953.

Картан Э. Геометрия римановых пространств, ОНТИ, 1936.

Специальная монография, широко освещающая предмет для подготовленного читателя.

---

[1] См. также литературу к главам VII (том 2) и XVIII.

Глава XVIII

ТОПОЛОГИЯ

§ 1. ПРЕДМЕТ ТОПОЛОГИИ

«*Прикосновение* составляет отличительную принадлежность тел и дает им название *геометрических*, когда в них удерживаем это свойство, не принимая в рассуждение все другие, существенные ли то будут или случайные».

Этими словами Н. И. Лобачевский начинает первую главу своего сочинения «Новые начала геометрии» [1].

Пояснив только что приведенные слова чертежом (рис. 1), Лобачевский продолжает: «Два тела A, B, касаясь друг друга, составляют одно геометрическое тело C. . . Обратно, всякое тело C произвольным сечением S разделяется на две части A, B».

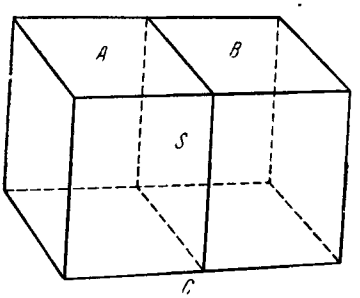


Рис. 1.

Эти понятия прикосновения, соседства, бесконечной близости, а также в некотором смысле двойственное им понятие рассечения тела — понятия, положенные Лобачевским в основу здания всей геометрии, и являются по существу основными, преимущественными понятиями топологии во всем том объеме, в котором мы теперь понимаем эту дисциплину. Поэтому правы современные комментаторы великого геометра, когда они говорят [2], что «Лобачевский делает первую в истории математических наук попытку исходить в построении геометрии от топологических свойств тел... Понятия поверхность, линия, точка определяются у Лобачевского в терминах сечений и прикосновений тел». Некоторое представление о разнообразии конкретного геометрического содержания, которое могут отражать понятия прикосновения и сечения тел, как их представлял себе Лобачевский, можно получить из прилагаемых чертежей (рис. 2), заимствованных из уже упомянутого его сочинения.

---

[1] Н. И. Лобачевский. Полн. собр. соч., т. II, Гостехиздат, 1949, стр. 168.

[2] Примечания к «Новым началам геометрии». Там же, стр. 465.

Всякое преобразование геометрической фигуры, при котором не разрушаются отношения прикосновения различных частей фигуры, называется *непрерывным*; если прикосновения не только не разрушаются, но и не возникают вновь, то преобразование называется *топологическим*. Следо-

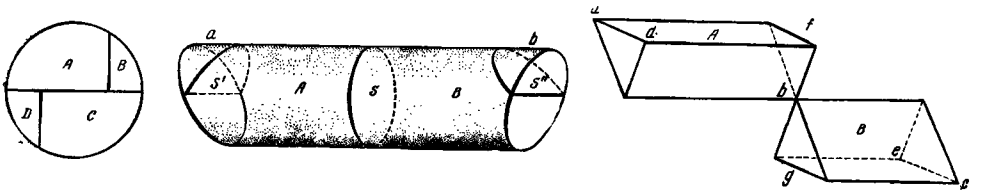


Рис. 2.

вательно, при топологическом преобразовании какой-либо фигуры, части этой фигуры, находящиеся в соприкосновении, остаются соприкасающимися, а части, не соприкасавшиеся, не могут стать соприкасающимися; короче говоря, при топологическом преобразовании не происходит ни разрывов, ни склеиваний. В частности, две различные точки не могут слиться в одну точку (в этом случае произошло бы новое прикосновение;

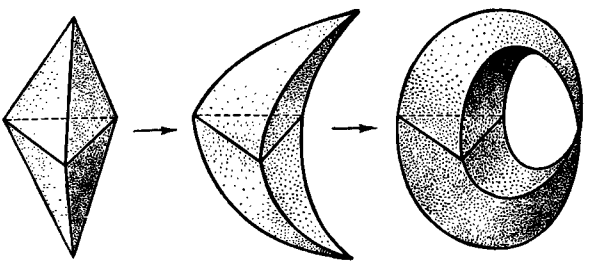


Рис. 3.

рис. 3). Поэтому топологическое преобразование всякой геометрической фигуры, рассматриваемой как множество образующих ее точек, есть преобразование не только непрерывное, но и взаимно однозначное: каждые две различные точки фигуры преобразуются в две различные точки. Таким образом, топологические преобразования являются взаимно однозначными и взаимно непрерывными.

Наглядно, топологическое преобразование какой-либо геометрической фигуры (линии, поверхности и т. п.) можно себе представить следующим образом.

Предположим, что наша фигура изготовлена из какого-нибудь гибкого и растяжимого материала, например из резины. Тогда можно подвергать ее всевозможным непрерывным деформациям, при которых она в одних своих частях будет растягиваться, в других — сжиматься и вообще будет всячески изменять свои размеры и свою форму. Например, придав замкнутой резиновой нити форму окружности, мы можем затем растянуть ее

в чрезвычайно вытянутый эллипс, можем придать ей форму правильного или неправильного многоугольника, а также формы весьма причудливых замкнутых кривых, некоторые из которых изображены на рис. 4. Но мы не можем посредством топологического преобразования превратить окружность в восьмерку (для этого пришлось бы склеить две различные точки окружности; рис. 5) или в отрезок (для этого пришлось бы склеить одну полуокружность с другой или, наоборот, разорвать окружность в какой-

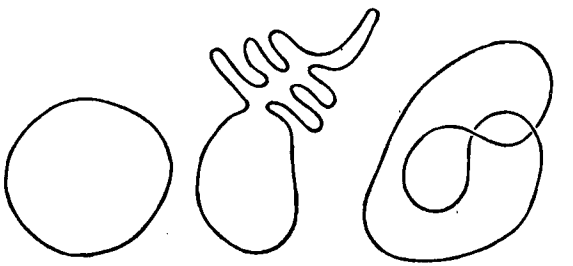


Рис. 4.

либо ее точке). Окружность есть простая замкнутая линия — она образует лишь одну петлю в отличие от восьмерки, которая образует две петли, или от трехлепестковой кривой (рис. 5), которая образует три петли. Свойство окружности быть простой замкнутой линией является свойством, сохраняющимся при произвольном ее топологическом преобразовании, или, как говорят, топологическим свойством.

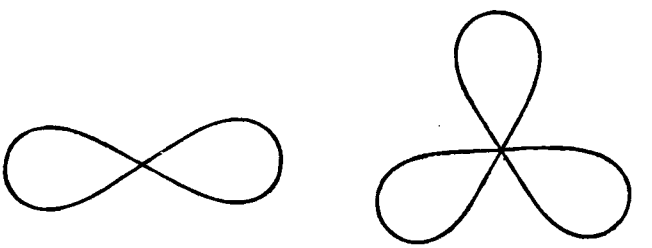


Рис. 5.

Если мы возьмем шаровую поверхность, которую можно представлять себе в виде тонкого резинового мячика, то мы снова посредством топологического преобразования можем чрезвычайно изменить ее форму (рис. 6). Но мы не сможем посредством топологического преобразования превратить нашу сферическую поверхность в квадрат или в кольцевидную поверхность (поверхность баранки или спасательного круга), называемую тором (рис. 7). В самом деле, поверхность сферы обладает следующими двумя свойствами, каждое из которых сохраняется при любом топологическом преобразовании. Первое свойство заключается в замкнутости нашей поверхности: у нее нет краев (а у квадрата есть край); второе свойство заключается в том, что всякая замкнутая линия на сферической

поверхности является, по выражению Лобачевского, ее сечением; если по данной замкнутой линии, начерченной на нашем резиновом мячике, сделать разрез, то поверхность распадется на две не связанные между со-

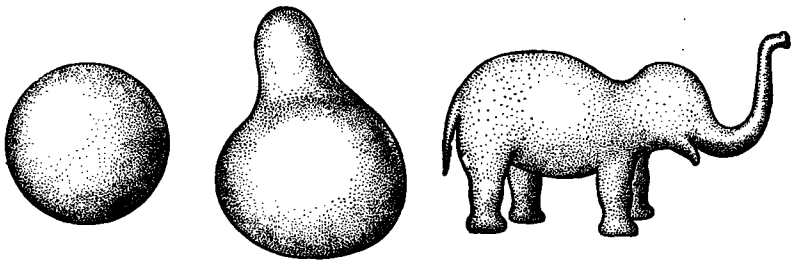


Рис. 6.

бой части. Тор этим свойством не обладает: если разрезать тор по его меридиану (рис. 7), то он не распадется на части, а превратится в поверхность, имеющую вид согнутой трубки (рис. 8), которую потом уже легко превратить топологическим преобразованием в цилиндр (разогнуть).

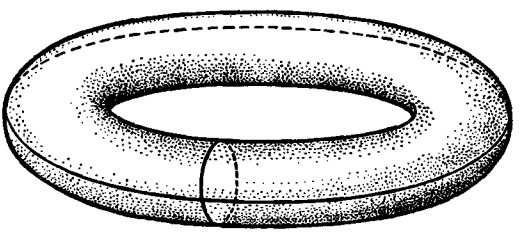


Рис. 7.

Итак, в отличие от сферы, на торе не всякая замкнутая линия является сечением. Поэтому сферическую поверхность нельзя топологическим преобразованием превратить в тор. Говорят, что сфера и тор суть топологически

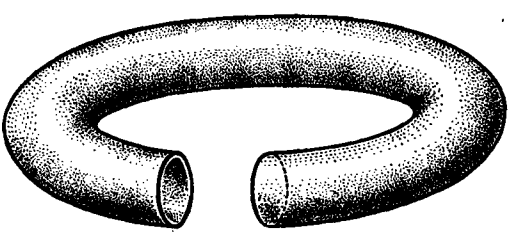


Рис. 8.

различные поверхности или поверхности, принадлежащие к различным топологическим типам, или, наконец, что эти поверхности не гомеоморфны между собой. Наоборот, сфера и эллипсоид и вообще любые ограниченные выпуклые поверхности принадлежат к одному и тому же топологическому типу, т. е. *гомеоморфны* между собой. Это значит, что они могут быть переведены одна в другую топологическим преобразованием.

§ 2. ПОВЕРХНОСТИ

Как уже упоминалось выше, всякое свойство геометрической фигуры, сохраняющееся при любом ее топологическом преобразовании, называется топологическим свойством. Топология изучает топологические свойства фигур; кроме того, она изучает топологические, а также любые непрерывные преобразования геометрических фигур.

Мы только что привели некоторые примеры топологических свойств. Так, топологическими свойствами являются: свойство замкнутости кривой или поверхности, свойство замкнутой линии быть простой замкнутой линией (т. е. образовывать лишь одну петлю), свойство замкнутой поверхности, состоящее в том, что всякая лежащая на ней замкнутая линия

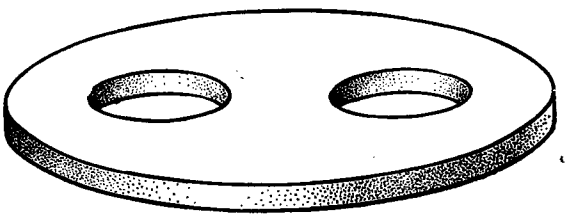


Рис. 9.

является сечением поверхности (этим свойством шаровая поверхность обладает, а кольцевидная нет), и др.

Наибольшее число замкнутых линий, которые можно провести на данной поверхности таким образом, чтобы эти линии не образовывали сечения, т. е. чтобы поверхность не распадалась на части, если по всем этим линиям сделать разрезы, называется *порядком связности* поверхности. Это число, пожалуй, дает нам самую важную информацию о топологическом устройстве поверхности. Мы видели, что для сферической поверхности оно равно нулю (всякая замкнутая линия на такой поверхности является сечением). На торе можно найти две замкнутые линии, которые в своей совокупности не образуют сечения: за одну из них можно принять любой меридиан, а за другую — параллель тора (рис. 7). Однако на торе невозможно провести три замкнутые линии, которые в своей совокупности не образовывали бы сечения; порядок связности тора равен двум. Порядок связности поверхности кренделя (рис. 9) равен четырем и т. д. Вообще возьмем сферическую поверхность и сделаем в ней $2p$ круглых отверстия (на рис. 10 изображен случай $p=3$). Эти отверстия распределим в p пар и к каждой паре отверстий приклеим (по краям) цилиндрическую трубку («ручку»). Получим сферу с p «ручками» или, как говорят, нормальную поверхность рода p. Порядок связности этой поверхности равен $2p$.

Все такие поверхности, по выражению Лобачевского, являются «сечениями» пространства: каждая из них разбивает пространство на две

области, внутреннюю и внешнюю, и является совместной границей этих двух областей. Это обстоятельство стоит в связи с другим, а именно с тем, что каждая из наших поверхностей имеет две стороны: внешнюю и внутреннюю (одну сторону можно покрасить в один цвет, а другую — в другой).

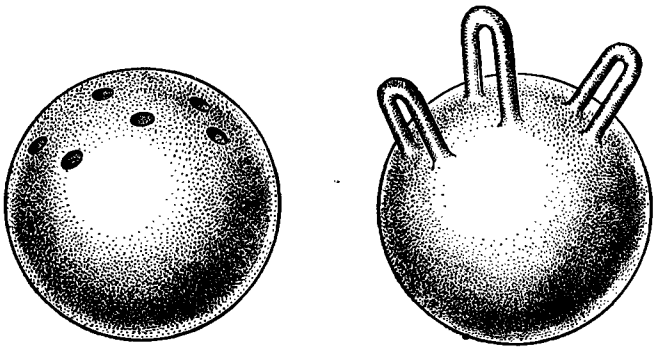


Рис. 10.

Однако наряду с этим существуют и так называемые односторонние поверхности, у которых нет этих двух раздельных сторон. Простейшей из них является всем известная «лента Мёбиуса», получающаяся, если взять прямоугольную полоску бумаги $ABCD$ и склеить две противоположные узкие стороны AB и CD прямоугольника $ABCD$ так, чтобы вершина A совпала с вершиной C, а вершина B с вершиной D. Полу-

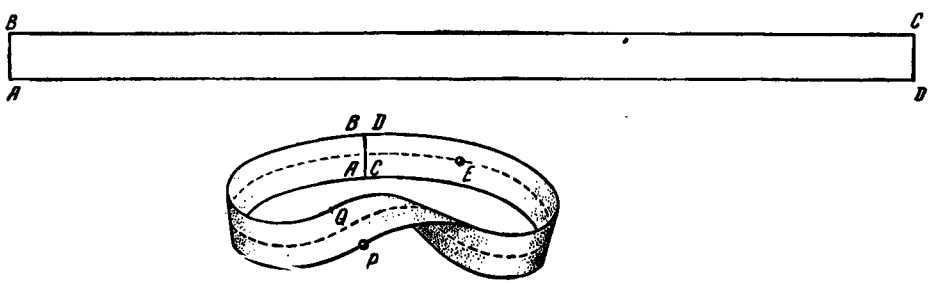


Рис. 11.

чится поверхность, изображенная на рис. 11, она и называется лентой или листом Мёбиуса. Нетрудно проверить, что на ней нет двух сторон, которые можно было бы выкрасить в разные цвета: идя по средней линии поверхности и начав свое движение, скажем, в точке E, мы, обойдя всю поверхность, придем в точку E, но уже с другой стороны поверхности, хотя и не перейдем при этом через ее край. Кстати, край поверхности Мёбиуса состоит из одной единственной замкнутой линии.

Возникает вопрос, существуют ли замкнутые односторонние поверхности, т. е. односторонние поверхности, не имеющие краев. Оказывается, что существуют, но такие поверхности, как бы мы их ни располагали в трех-

мерном пространстве, всегда имеют самопересечения. Типичный пример замкнутой односторонней поверхности изображен на рис. 12 — это так называемый «односторонний тор» или поверхность Клейна. Если, не боясь самопересечений, мыслить себе два экземпляра листа Мёбиуса склеенными по их краям (край листа Мёбиуса, как уже упоминалось, состоит лишь из одного контура), то получится поверхность Клейна.

Теперь мы можем сформулировать основную теорему топологии поверхностей в применении к двусторонним поверхностям: всякая замкнутая двусторонняя поверхность гомеоморфна некоторой *нормальной поверхности* рода p, т. е. «сфере с p ручками»; две замкнутые двусторонние

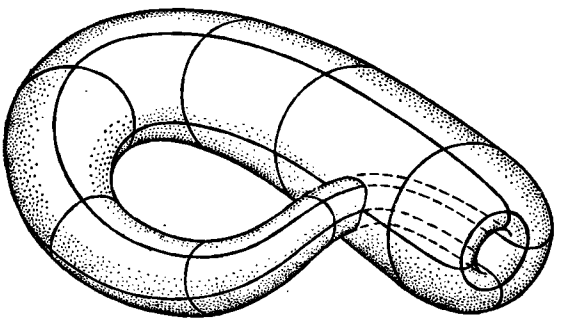


Рис. 12.

поверхности гомеоморфны тогда и только тогда, когда у них один и тот же *род* p (тот же порядок связности $2p$), т. е. когда они гомеоморфны сфере с одним и тем же числом ручек p.

Для односторонних поверхностей также имеются «нормальные формы», аналогичные нормальным формам двусторонних поверхностей рода p, но их труднее себе представить. Для этого надо взять сферу, проделать в ней p круглых отверстий и заклеить каждое из них поверхностью Мёбиуса, склеив край этой поверхности с краем соответствующего отверстия. Трудность, возникающая при попытке представить себе такое склеивание, происходит от того, что осуществить его физически нельзя: при попытке такого склеивания возникнут как раз те самопересечения поверхности, которые неизбежны при всяком осуществлении односторонней замкнутой поверхности в виде пространственной модели.

Не следует думать, что замкнутые односторонние поверхности относятся к области математических курьезов, не связанных с серьезными задачами науки. Чтобы убедиться в ошибочности такого мнения, достаточно вспомнить, что одним из основных достижений геометрической мысли явилось создание так называемой проективной геометрии, элементы которой входят в настоящее время в курсы геометрии университетов и педагогических институтов. Практические истоки проективной геометрии лежат в теории перспективы, возникшей еще

в эпоху Возрождения (Леонардо да Винчи) в связи с потребностями архитектуры, живописи и технического проектирования. К XVI—XVII вв. относится открытие первых теорем проективной геометрии. Возникнув, таким образом, в связи с вполне определенными практическими потребностями, проективная геометрия явилась в своем полном развитии одним из значительнейших идейно-теоретических обобщений геометрии. На ее почве, в частности, впервые была до конца понята неэвклидова геометрия Лобачевского [1].

Переход от обычной плоскости, как ее изучает элементарная геометрия, к проективной плоскости заключается в пополнении плоскости новыми абстрактными элементами, так называемыми несобственными или «бесконечно удаленными» точками. Только после такого пополнения операция проектирования одной плоскости на другую (например, проектирования на экран посредством проекционного фонаря) становится взаимно однозначным преобразованием одной плоскости в другую. Пополнение плоскости несобственными точками, которому в аналитической геометрии соответствует переход от обыкновенных декартовых координат к однородным координатам, происходит следующим образом. Каждая прямая дополняется одной единственной несобственной («бесконечно удаленной») точкой, причем две прямые тогда и только тогда имеют одну и ту же несобственную точку, когда они параллельны. Пополненная единственной бесконечно удаленной точкой прямая становится замкнутой линией, а совокупность бесконечно удаленных точек всевозможных прямых по определению образует несобственную или бесконечно удаленную прямую.

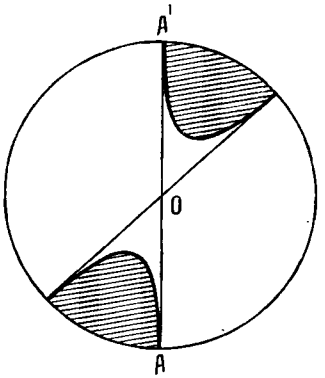


Рис. 13.

Так как параллельные прямые имеют общую бесконечно удаленную точку, то для того, чтобы представить себе весь процесс пополнения плоскости несобственными точками, достаточно рассмотреть прямые, проходящие через одну какую-нибудь точку плоскости, например через начало координат O (рис. 13). Несобственные точки этих прямых уже исчерпают все вообще несобственные точки проективной плоскости (так как каждая прямая имеет ту же несобственную точку, что и параллельная ей прямая, проходящая через точку O). Поэтому мы получим «модель» проективной плоскости, рассматривая ее как круг «бесконечно большого» радиуса с центром в O и считая, что любая пара диаметрально противоположных точек A, A' окружности этого круга должна быть склеена в одну «беско-

[1] См., например, главу XVII, § 6, или книжку П. С. Александрова «Что такое неэвклидова геометрия» (М., 1951).

нечно удаленную» точку прямой AA'. Сама окружность нашего круга превращается при этом в бесконечно удаленную прямую, но при этом надо твердо помнить, что любые две диаметрально противоположные точки этой окружности должны мыслиться как отождествленные между собой. Отсюда сразу видно, что проективная плоскость есть замкнутая поверхность, никаких краев у нее нет.

Если мы рассмотрим на проективной плоскости кривую второго порядка, которую рисуем в виде гиперболы (см. рис. 13), то ясно, что эта гипербола на проективной плоскости представляет собой замкнутую кривую (лишь разрезанную на две ветви бесконечно удаленной прямой). Имея в виду, что диаметрально противоположные точки окружности нашего основного круга склеены между собой, мы без труда убеждаемся, что заштрихованная на рис. 13 внутренность нашей гиперболы гомеоморфна внутренности обыкновенного круга, а дополнительная, незаштрихованная часть проективной плоскости гомеоморфна ленте Мёбиуса. Таким образом, проективная плоскость с топологической точки зрения есть результат склеивания круга (в нашем случае — внутренности гиперболы) с листом Мёбиуса по их краю. Отсюда следует, что проективная плоскость, т. е. основной объект изучения плоской проективной геометрии, есть замкнутая односторонняя поверхность.

Пример проективной плоскости, кроме своего большого принципиально геометрического значения, интересен еще и потому, что на нем рельефно выделяется одна особенность современного геометрического мышления, сформировавшегося на основе открытий Лобачевского. Геометрическое мышление всегда, по самому характеру понятия геометрической фигуры, было абстрактным. Теперь оно поднимается на новую ступень абстракции, проявляющуюся в нашем случае в пополнении обычной плоскости новыми абстрактными элементами — несобственными точками. Конечно, и эти абстрактные элементы отражают реальную действительность (каждая «несобственная точка» есть не что иное, как абстракция пучка параллельных прямых), но входят они в наши рассмотрения как отвлеченные геометрические элементы, которые мы лишь несовершенно можем себе представить в виде результата (физически не осуществимого) «склеивания» диаметрально противоположных точек окружности некоторого круга. Аналогичные абстрактные построения имеют очень большое значение во всей современной топологии, особенно при переходе от поверхностей к многообразиям трех и более измерений.

§ 3. МНОГООБРАЗИЯ

Рассмотрим следующий простой прибор, иногда называемый двойным плоским маятником (рис. 14). Он состоит из двух стержней OA и AB, скрепленных шарниром в точке A; точка O остается неподвижной, стержень OA свободно вращается в постоянной плоскости вокруг точки O, а стержень AB свободно вращается в той же плоскости вокруг точки A.

Каждое из возможных положений нашей системы вполне определяется, если задать угол φ и угол ψ, которые образуют стержни OA и AB с каким-нибудь постоянным направлением в плоскости, например с положительным направлением оси абсцисс. Эти два угла, изменяющиеся от 0 до 2π мы можем рассматривать как «географические координаты» точки на торе, отсчитываемые соответственно от «экватора» тора и одного из его «меридианов» (рис. 15).

Таким образом, мы можем сказать, что многообразие всех возможных состояний нашей механической системы есть многообразие двух измерений, а именно — тор. Заменяя каждый из углов φ, ψ соответствующей ему точкой некоторой окружности, на которой дана начальная точка и направление отсчета дуг (рис. 16), мы можем также сказать, что каждое

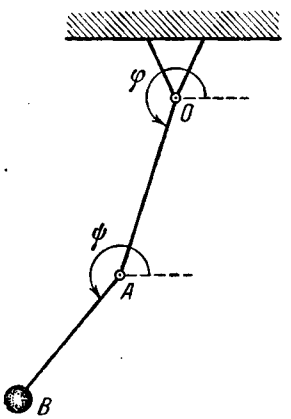


Рис. 14.

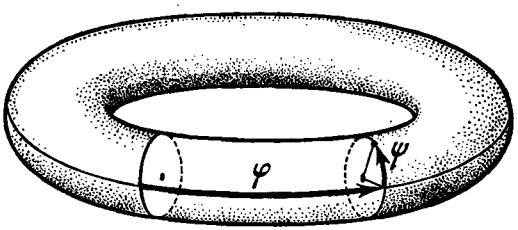


Рис. 15.

возможное состояние нашей механической системы будет вполне охарактеризовано, если будет задано по точке на каждой из двух окружностей (на одной из них отсчитывается широта φ, а на другой — долгота ψ). Другими словами, совершенно так же, как в аналитической геометрии, мы отождествляем точку плоскости с парой чисел — ее координат, так и в нашем случае мы можем отождествить точку тора (а значит и любое положение нашего маятника) с парой ее географических координат, т. е. с парой точек, из которых одна лежит на одной, а другая на другой окружности. Это положение вещей характеризуют, говоря, что многообразие всех возможных состояний нашего двойного плоского маятника, т. е. тор, есть топологическое произведение двух окружностей.

Видоизменим теперь наш прибор следующим образом. Пусть он попрежнему состоит из двух стержней OA и AB, причем стержень OA может свободно вращаться в определенной плоскости вокруг точки O, тогда как стержень AB теперь скреплен со стержнем OA шаровым шарниром в точке A, так что при данном положении этой точки он может свободно вращаться вокруг нее уже в пространстве, принимая направление любого исходящего из точки A луча. Теперь положение нашей системы задается тремя параметрами, из которых первый есть попрежнему угол φ, образуемый стержнем OA с положительным направлением оси абсцисс, а вторые

два определяют направление стержня AB в пространстве. Последнее направление может быть определено, например, заданием той точки B' единичной сферы с центром в начале координат O, в которой пересекает эту сферу радиус OB', параллельный стержню AB, или заданием на сфере двух географических координат точки B'. Таким образом, многообразие всех положений нашей новой шарнирной системы есть некоторое трехмерное многообразие, и читатель легко поймет, что его можно трактовать как топологическое произведение окружности и сферы. Это многообразие замкнуто, т. е. оно не имеет краев, поэтому его невозможно реализовать в виде фигуры, лежащей в трехмерном пространстве. Для того чтобы тем не менее представить себе это многообразие сколько-нибудь наглядно, рассмотрим часть пространства, заключенную между двумя концентрическими сферами. Каждый луч, идущий из общего центра этих сфер, прокалывает их в двух точках. Если считать каждую пару таких точек отождествленной (склеенной в одну точку), то мы и получим трехмерное многообразие, являющееся топологическим произведением сферы на окружность.

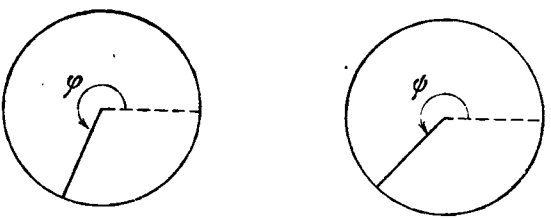


Рис. 16.

Мы можем подвергнуть наш шарнирный прибор еще дальнейшему усложнению, не только скрепив шаровым шарниром стержни OA и AB в точке A, но предположив, кроме того, что и стержень OA может свободно вращаться в пространстве вокруг точки O. Множество возможных положений полученной системы будет уже четырехмерным замкнутым многообразием, а именно — произведением двух сфер.

Мы видим, таким образом, что уже простейшие механические рассмотрения (кинематические) приводят к топологическим многообразиям и притом трех и более измерений. При реальном, более подробном рассмотрении механических проблем еще большее значение имеют многообразия (вообще говоря, также многомерные), являющиеся так называемыми *фазовыми пространствами* динамических систем, где принимаются в расчет не только конфигурации, которые может иметь данная механическая система, но и скорости, с которыми движутся различные составляющие ее точки. Ограничимся одним из простейших примеров. Пусть мы имеем точку, способную двигаться по окружности с произвольной скоростью. Каждое состояние этой системы определяется двумя данными: положением точки на окружности и скоростью ее в данный момент. Многообразием состояний (фазовым пространством) данной механической системы является, очевидно, бесконечный цилиндр (произведение окружности на прямую).

Число измерений фазового пространства растет вместе с ростом числа степеней свободы данной системы. Многие динамические характеристики

той или иной механической системы находят свое выражение в топологических свойствах ее фазового пространства. Так, например, каждому периодическому движению данной системы соответствует замкнутая линия в ее фазовом пространстве.

Изучение фазовых пространств динамических систем, поставленное на очередь различными проблемами механики, физики и астрономии (небесная механика, космогония), привлекло внимание математиков к топологии многомерных многообразий. Именно в связи с этими проблемами великий французский математик Пуанкаре предпринял в девяностых годах прошлого века систематическое построение топологии многообразий, применив при этом так называемый комбинаторный метод, являющийся с тех пор одним из основных методов топологии.

§ 4. КОМБИНАТОРНЫЙ МЕТОД

Исторически первой теоремой, относящейся к топологии, является теорема или формула Эйлера (повидимому, известная еще Декарту). Она заключается в следующем. Возьмем поверхность произвольного выпуклого многогранника. Обозначим через α_0 число ее вершин, через α_1 число ребер и через α_2 число граней; тогда имеет место следующее соотношение, которое и известно под названием *формулы Эйлера:*

$$\alpha_0 - \alpha_1 + \alpha_2 = 2. \qquad (1)$$

Эта геометрическая теорема именно потому относится к топологии, что наша формула, очевидно, останется верной, если мы подвергнем рассматриваемый выпуклый многогранник произвольному топологическому преобразованию. При таком преобразовании ребра, вообще говоря, перестанут быть прямолинейными, грани — плоскими, поверхность многогранника перейдет в кривую поверхность, но соотношение (1) между числом вершин и числами теперь уже кривых ребер и граней останется в силе. Наиболее важен случай, когда все грани треугольные и мы имеем так называемую *триангуляцию* (разбиение нашей поверхности на треугольники — прямолинейные или криволинейные). К этому случаю легко приводится общий случай любых многоугольных граней: достаточно разбить эти грани на треугольники (например, проведением диагоналей из какой-либо вершины данной грани). Таким образом, мы можем ограничиться именно случаем триангуляций. Комбинаторный метод в топологии поверхностей и заключается в том, что изучение этой поверхности заменяется изучением ее триангуляций, причем нас интересуют, конечно, только те свойства триангуляций, которые не зависят от случайного выбора той или иной триангуляции, а, будучи общими для всех триангуляций данной поверхности, выражают некоторое свойство самой поверхности.

Формула Эйлера приводит нас к одному из таких свойств, и мы им сейчас займемся несколько подробнее. Левая часть формулы Эйлера, т. е.

выражение $\alpha_0 - \alpha_1 + \alpha_2$, где α_0 — число вершин, α_1 — число ребер и α_2 — число треугольников в данной триангуляции, называется *эйлеровой характеристикой* этой триангуляции. Теорема Эйлера утверждает, что для всех триангуляций поверхности, гомеоморфной сфере, эйлерова характеристика равна двум. Оказывается, что для всякой поверхности (а не только поверхности, гомеоморфной сфере) все триангуляции этой поверхности имеют одну и ту же эйлерову характеристику.

Легко сообразить, каково будет значение эйлеровой характеристики для различных поверхностей. Прежде всего, для цилиндрической поверхности оно равно нулю. В самом деле, удалив из какой-либо триангуляции сферы два не соприкасающиеся между собой треугольника, но сохраняя границы этих треугольников, мы, очевидно, получим триангуляцию поверхности, гомеоморфной боковой поверхности цилиндра. При этом число вершин и ребер осталось прежним, а число треугольников уменьшилось на 2, поэтому эйлерова характеристика полученной триангуляции равна нулю. Возьмем теперь поверхность, полученную из триангуляции сферы посредством удаления $2p$ треугольников этой триангуляции, попарно не соприкасающихся (т. е. не имеющих ни общих вершин, ни общих сторон) [1]. При этом эйлерова характеристика уменьшается на $2p$ единиц. Легко убедиться в том, что при заклеивании каждой пары образовавшихся в поверхности сферы отверстий цилиндрической трубкой эйлерова характеристика не изменится. Это происходит оттого, что характеристика самой приклеиваемой трубки равна, как мы видели, нулю, а на краях этой трубки число вершин равно числу ребер. Итак, замкнутая двусторонняя поверхность рода p имеет эйлерову характеристику $2-2p$ (факт, впервые доказанный французским адмиралом де Жонкьером).

Приведем еще одно важное свойство триангуляций, удовлетворяющее так называемому условию топологической инвариантности (т. е. принадлежащее каждой триангуляции данной поверхности, если оно принадлежит хотя бы одной из них). Это — свойство *ориентируемости*. Прежде чем его формулировать, заметим, что каждый треугольник можно ориентировать, т. е. снабдить его границу определенным направлением обхода. Каждая из двух возможных ориентаций треугольника задается определенным порядком следования его вершин [2]. Предположим теперь, что на какой-нибудь поверхности даны два треугольника, примыкающие

[1] Для того чтобы это можно было сделать, надо только, чтобы рассматриваемая триангуляция была достаточно «мелкой». Этого всегда можно достигнуть надлежащим подразделением любой триангуляции.

[2] При этом легко видеть, что два порядка вершин тогда и только тогда определяют одну и ту же ориентацию (одно и то же направление обхода), когда они переходят друг в друга посредством «четной» перестановки. Так, (ABC), (BCA), (CAB) определяют одну, а (BAC), (ACB), (CBA) другую ориентацию треугольника. (О четных и нечетных перестановках см., например, главу XX, § 3.)

друг к другу по общей стороне и не имеющие других общих точек (рис. 17). Две ориентации этих треугольников называются согласованными, если они порождают на общей стороне треугольников противоположные направления. (На плоскости или на любой другой двусторонней поверхности это означает, что оба треугольника — если смотреть на них с одной стороны поверхности — обходятся в одном направлении, т. е. либо оба против, либо оба по часовой стрелке.) Триангуляция данной замкнутой поверхности называется ориентируемой, если ориентации всех входящих в эту триангуляцию треугольников можно выбрать так, что любые два прилегающие по общей стороне треугольника окажутся ориентированными согласованно. Имеет место такой факт: всякая триангуляция

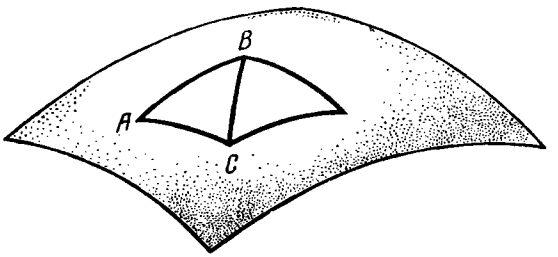


Рис. 17.

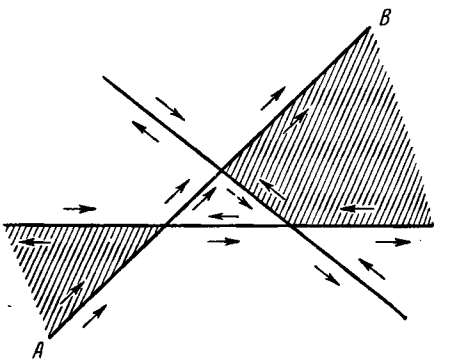


Рис. 18.

двусторонней поверхности ориентируема, всякая триангуляция односторонней поверхности неориентируема. Поэтому двусторонние поверхности называются также ориентируемыми, а односторонние — неориентируемыми. Взяв произвольную триангуляцию листа Мёбиуса, читатель без труда убедится в ее неориентируемости. Чтобы получить простейшую триангуляцию проективной плоскости, надо на ней провести какие-либо три прямые, не проходящие через одну точку (рис. 18). Они разобьют проективную плоскость на четыре треугольника, из которых один лежит в конечной части плоскости, а три других рассечены, каждый на две части, бесконечно удаленной прямой. На рис. 18 один из этих трех уходящих в бесконечность треугольников заштрихован. На этом же чертеже видно, что, пытаясь придать всем четырем треугольникам согласованные ориентации, мы неизбежно приходим к неудаче. В частности, при сделанном на рис. 18 выборе ориентаций наших четырех треугольников, мы в качестве алгебраической суммы их границ вместо нуля, который получился бы при согласованных ориентациях, получаем дважды взятую прямую AB.

Эйлерова характеристика и свойство ориентируемости или неориентируемости замкнутых поверхностей дают нам, как говорят, полную систему топологических инвариантов замкнутых поверхностей. Смысл

этого высказывания заключается в том, что две поверхности тогда и только тогда гомеоморфны, когда триангуляции этих поверхностей, во-первых, имеют одну и ту же эйлерову характеристику и, во-вторых, обе ориентируемы или обе неориентируемы.

Комбинаторный метод применяется не только к изучению поверхностей (двумерных многообразий), но и многообразий любого числа измерений. Но, например, в случае трехмерных многообразий роль обычных триангуляций играют уже разбиения на тетраэдры. Их называют трехмерными триангуляциями или симплициальными разбиениями многообразия. Под эйлеровой характеристикой трехмерной триангуляции понимается число $\alpha_0 - \alpha_1 + \alpha_2 - \alpha_3$, где α_i, $i = 0, 1, 2, 3$, есть число i-мерных элементов этой триангуляции (т. е. α_0 — число вершин, α_1 — число ребер, α_2 — число двумерных граней, α_3 — число тетраэдров). При числе измерений $n > 3$ многообразие разбивают на n-мерные симплексы, т. е. простейшие выпуклые n-мерные многогранники, аналогичные треугольникам ($n = 2$) и тетраэдрам ($n = 3$). Симплексы, на которые разбито n-мерное многообразие, и их грани образуют n-мерную триангуляцию этого многообразия. Попрежнему можно говорить об эйлеровой характеристике, понимая под ней сумму $\sum_{i=0}^{n} (-1)^i \alpha_i$, где α_i — число входящих в данную триангуляцию i-мерных элементов ($i = 0, 1, 2, \ldots, n$), и попрежнему эйлерова характеристика имеет одно и то же значение для всех триангуляций данного n-мерного многообразия (и всех многообразий, гомеоморфных данному), т. е. является топологическим инвариантом. Но о полной системе инвариантов (в том смысле, в каком она для поверхностей дается эйлеровой характеристикой и ориентируемостью) мы даже для трехмерных многообразий при настоящем уровне наших знаний не можем и мечтать!

Значение комбинаторного метода в современной топологии очень велико. Этот метод открывает возможность применять некоторые алгебраические приемы в решении топологических задач. Возможность такого алгебраического подхода внимательный читатель мог уже усмотреть выше, когда мы говорили об алгебраической сумме границ ориентированных треугольников в триангуляции проективной плоскости. В самом деле, если треугольник ориентирован, т. е. снабжен определенным направлением обхода, то за его границу естественно принять совокупность его сторон, взятых каждая также с определенным направлением — тем, которое порождается имеющимся обходом треугольников.

Рассмотрим теперь все треугольники T_i^2, $i = 1, 2, \ldots, \alpha_2$, входящие в данную триангуляцию некоторой поверхности. Каждому из них можно придать две ориентации; треугольник T_i^2, с одной из двух

возможных ориентаций обозначим через t_i^2, а тот же треугольник с другой (противоположной) ориентацией — через t_i^2. Точно так же каждый из одномерных элементов (ребер) T_k^1 ($k=1, 2, \ldots, \alpha_1$), входящих в данную триангуляцию, можно ориентировать, т. е. снабдить его одним из двух возможных направлений. Отрезок T_k^1 с одной из ориентаций обозначим через t_k^1, с другой — через t_k^1. Тогда, если сторонами треугольника T_i^2 являются отрезки T_1^1, T_2^1, T_3^1, то границей ориентированного треугольника t_i^2 является совокупность тех же сторон, но взятых с определенными направлениями, т. е. граница состоит из направленных отрезков $\varepsilon_1 t_1^1$, $\varepsilon_2 t_2^1$, $\varepsilon_3 t_3^1$; здесь $\varepsilon_i = 1$, если это направление для ребра T_i^1 совпадает с его собственным направлением t_i^1 и $\varepsilon_i = -1$ в противном случае. Границу t_i^2 обозначают через Δt_i^2. Как видим, $\Delta t_i^2 = \sum \varepsilon_k t_k^1$, причем эту сумму можно представлять себе распространенной на все ребра нашей триангуляции, считая, что коэффициенты ε_k при отрезках, не входящих в границу треугольника t_i^2, равны нулю.

Становится естественным рассматривать вообще суммы вида $x^1 = \sum a_k t_k^1$, распространенные по всем ребрам данной триангуляции[1]. Геометрический смысл таких сумм очень прост: ведь каждое слагаемое суммы есть некоторый отрезок, входящий в нашу триангуляцию, взятый с определенным направлением и определенным коэффициентом (определенной «кратностью»). Вся же написанная алгебраическая сумма выражает составленный из отрезков путь, в котором каждый отрезок считается столько раз, сколько указывает стоящий при нем коэффициент. Например, если мы сначала проходим многоугольник $ABCDEF$ (рис. 19) в показанном стрелкой направлении, потом переходим по отрезку AA' на многоугольник $A'B'C'D'E'F'$, проходимый в указанном на нем направлении, затем возвращаемся по отрезку $A'A$ назад и проходим многоугольник $ABCDEF$ снова в том же, что и раньше, направлении, то получаем сумму, в которую отрезки AB, BC, CD, DE, EF, FA войдут с коэффициентом 2, отрезки $A'B'$, $B'C'$, $C'D'$, $D'E'$, $E'F'$, $F'A'$ — с коэффициентом 1, а отрезок AA' вовсе не войдет (он будет иметь коэффициент нуль, так как он оказывается пройденным два раза в двух противоположных направлениях).

Суммы вида $x^1 = \sum a_k t_k^1$ называются *одномерными цепями* данной триангуляции. С алгебраической точки зрения они представляют собой линейные формы (однородные многочлены первой степени); их можно складывать и вычитать, а также умножать на любое целое число по обычным правилам алгебры. Среди одномерных цепей особенно важными являются так называемые одномерные циклы. Геометрически они соот-

[1] Коэффициенты a_k будем считать целыми числами.

ветствуют замкнутым путям (именно о таком пути была только что речь в связи с рис. 19).

Для чисто алгебраического определения цикла условимся из двух вершин направленного отрезка $\overrightarrow{AB}$ считать конечную вершину B входящей в границу отрезка $\overrightarrow{AB}$ со знаком плюс (с коэффициентом $+1$), а начальную вершину A — со знаком минус (с коэффициентом -1). Тогда *границу* отрезка $\overrightarrow{AB}$ можно записать в виде $\Delta(\overrightarrow{AB}) = B - A$.

Приняв такое соглашение, мы непосредственно замечаем, что сумма границ отрезков $\overrightarrow{AB}$, $\overrightarrow{BC}$, $\overrightarrow{CD}$, ..., $\overrightarrow{FA}$, образующих замкнутый (в обычном смысле слова) путь, равна нулю. Это делает естественным общее определение одномерного *цикла*, как такой одномерной цепи $z^1 = \sum_k a_k t^1_k$, сумма границ звеньев которой, т. е. сумма $\sum_k a_k \Delta t^1_k$ равна нулю. Легко проверить, что сумма двух циклов есть цикл. Умножая цикл как алгебраическое выражение на какое-либо целое число, получим опять цикл. Это позволяет говорить о линейных комбинациях циклов z^1_1, z^1_2, ..., z^1_s, т. е. о циклах вида $z = \sum_{\nu=1}^{s} c_\nu z^1_\nu$, где c_ν — целые числа.

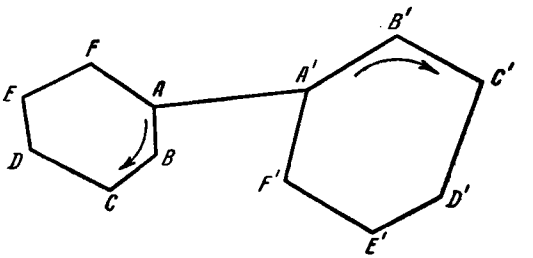


Рис. 19.

Аналогично понятию одномерной цепи данной триангуляции можно говорить и о двумерных цепях этой триангуляции, т. е. о выражениях вида $x^2 = \sum_i a_i t^2_i$, где t^2_i — ориентированные треугольники данной триангуляции. Так как граница каждого ориентированного треугольника есть одномерный цикл, то циклом является и цепь $\sum_i a_i \Delta t^2_i$. Именно этот цикл считается границей Δx^2 цепи $x^2 = \sum_i a_i t^2_i$.

Понятие *границы цепи* позволяет далее сформулировать понятие гомологии: одномерный цикл z^1 данной триангуляции называется *гомологичным нулю* в этой триангуляции, если он является границей некоторой двумерной цепи этой триангуляции. Во всякой триангуляции замкнутой выпуклой поверхности и вообще любой поверхности, гомеоморфной сфере, всякий одномерный цикл гомологичен нулю; геометрически это совершенно ясно: каждый замкнутый многоугольник на выпуклой поверхности является границей некоторого куска этой поверхности. Не то на торе: меридиан тора, так же как и его экватор, не является границей какого-либо куска этой поверхности. Если взять какую-либо триангуляцию тора, то в ней найдутся циклы, аналогичные

меридиану и экватору тора, и эти циклы не будут гомологичны нулю.

Совершенно новое явление мы наблюдаем на построенной выше триангуляции проективной плоскости. Если рассматривать прямую, например прямую AB (см. рис. 18), как цикл этой триангуляции, то этот цикл не гомологичен на ней нулю. Однако та же прямая, взятая с коэффициентом 2, уже оказывается гомологичной нулю. Таким образом, привлечение коэффициентов, отличных от ± 1, при определении цепи, которое кажется сначала формальным и ненужным, позволяет уловить важные геометрические свойства поверхностей и вообще многообразий. В данном случае это так называемое свойство кручения, заключающееся в существовании циклов, которые в данном многообразии не гомологичны нулю (не ограничивают никакого куска поверхности), но становятся гомологичными нулю, если снабдить их некоторым целочисленным коэффициентом.

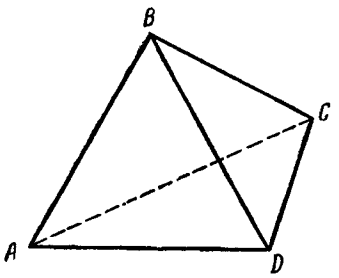


Рис. 20.

В связи со сказанным введем, наконец, чрезвычайно важное понятие *гомологической независимости* циклов. Циклы $z_1, \ldots, z_s$ называются гомологически независимыми в данной триангуляции, если никакая их линейная комбинация $\sum c_i z_i$, в которой хотя бы один коэффициент c_i отличен от нуля, не гомологична нулю в этой триангуляции. Примерами гомологически независимых циклов на торе могут служить какой-либо меридиан и экватор тора, рассматриваемые как циклы некоторой триангуляции тора.

Основные понятия всей комбинаторной топологии — понятия границы, цикла, гомологии — определены нами для одномерных образований, но они дословно переносятся и на любое число измерений. Так, например, двумерная цепь $z^2 = \sum a_i t_i^2$ называется циклом, если ее граница $\Delta z^2 = \sum a_i \Delta t_i^2$ равна нулю. Трехмерная цепь есть выражение вида $x^3 = \sum a_k t_k^3$, где t_k^3 — ориентированные трехмерные симплексы (тетраэдры).

Как и в случае треугольника, ориентация трехмерного симплекса (тетраэдра) задается определенным порядком его вершин, причем два порядка вершин, переходящие друг в друга при четной перестановке, определяют одну и ту же ориентацию. Граница трехмерного ориентированного симплекса $t^3 = (ABCD)$ есть двумерная цепь (цикл) $\Delta t^3 = (BCD) - (ACD) + (ABD) - (ABC)$ (рис. 20). Граница трехмерной цепи определяется как сумма границ ее симплексов, взятых с теми коэффициентами, с которыми эти симплексы входят в данную цепь. Читатель легко проверит, что граница любой трехмерной цепи есть

двумерный цикл (это достаточно доказать для границы одного трехмерного симплекса). Мы говорим, что двумерный цикл гомологичен нулю в данном многообразии, если он является границей некоторой трехмерной цепи этого многообразия. И так далее. Заметим, что из данного выше определения ориентируемых и неориентируемых триангуляций[1] сразу следует, что во всякой ориентируемой триангуляции имеются (в случае поверхностей — двумерные) циклы, отличные от нуля, а в неориентируемых триангуляциях таких циклов нет; этот результат также непосредственно обобщается на любое число измерений.

Введенные понятия позволяют определить порядок одномерной, двумерной и т. д. связности данных многообразий произвольного числа измерений. Максимальное число имеющихся в какой-либо триангуляции данного многообразия гомологически независимых одномерных, двумерных и т. д. циклов не зависит от выбора триангуляций данного многообразия и называется его *порядком связности*, или числом Бетти (соответствующей размерности).

Одномерное число Бетти замкнутой ориентируемой поверхности рода p равно $2p$ (т. е. порядку связности поверхности, как он определялся в § 2). Одномерное число Бетти проективной плоскости равно 0. (Там всякий цикл, не уходящий в бесконечность, ограничивает часть плоскости, т. е. гомологичен нулю, а цикл, уходящий в бесконечность, например проективная прямая, оказывается гомологичным нулю, если взять его дважды.) Двумерное число Бетти всякой неориентируемой поверхности равно нулю (на такой поверхности нет ни одного отличного от нуля двумерного цикла).

Двумерное число Бетти всякой ориентируемой поверхности равно 1. Действительно, если ориентировать надлежащим образом все треугольники какой-нибудь триангуляции ориентируемой поверхности, получим цикл (так называемый основной цикл поверхности). Нетрудно проследить, что всякий двумерный цикл получается из основного цикла умножением его на какое-нибудь целое число. Эти результаты непосредственно обобщаются и на n-мерные многообразия. Заметим еще, что нульмерное число Бетти связного (т. е. не распадающегося на куски) многообразия считается равным 1.

Числа Бетти различных измерений связаны с эйлеровой характеристикой многообразия замечательной формулой, доказанной Пуанкаре и обобщающей теорему Эйлера. Эта формула, известная под названием *формулы Эйлера—Пуанкаре*, имеет следующий простой вид:

$$\sum_{r=0}^{n} (-1)^r \alpha_r = \sum_{r=0}^{n} (-1)^r p_r.$$

[1] Это определение, данное выше для триангуляций поверхностей, переносится и на случай триангуляций многообразий любого числа измерений.

Здесь слева стоит эйлерова характеристика произвольной триангуляции данного многообразия, а числа p_r справа суть числа Бетти различных размерностей r этого многообразия. В частности, для ориентируемых поверхностей имеем, как мы только что видели, $p_0 = p_2 = 1$, $p_1 = 2p$, где p — род поверхности. Это и дает нам теорему Эйлера для ориентируемых поверхностей

$$\alpha_0 - \alpha_1 + \alpha_2 = 2 - 2p.$$

§ 5. ВЕКТОРНЫЕ ПОЛЯ

Рассмотрим простейшее дифференциальное уравнение

$$\frac{dy}{dx} = F(x, y), \tag{2}$$

заданное в данной плоской области G. Его геометрический смысл заключается в том, что в каждой точке (x, y) области G определено направление, угловой коэффициент которого равен $F(x, y)$, где $F(x, y)$ — некоторая непрерывная функция точки (x, y). Как говорят, в области G задано непрерывное поле направлений; мы можем его легко превратить в непрерывное векторное поле, беря, например, в каждом из заданных направлений вектор единичной длины. Задача интегрирования дифференциального уравнения (2) заключается в том, чтобы разложить, если это возможно, данную плоскую область на попарно не пересекающиеся кривые («интегральные кривые» уравнения) таким образом, чтобы в каждой точке области заданное в ней направление было направлением касательной к единственной интегральной кривой, проходящей через эту точку.

Рассмотрим, например, уравнение

$$\frac{dy}{dx} = \frac{y}{x}.$$

В каждой точке $M(x, y)$ на плоскости отнесенное ей направление есть очевидно направление луча $\overrightarrow{OM}$ (где O — начало координат). Интегральные кривые суть прямые, проходящие через точку O. Через каждую отличную от O точку плоскости проходит единственная интегральная кривая. Что же касается начала координат, то это особая точка данного дифференциального уравнения (так называемый «узел»), через нее проходят все интегральные кривые.

Если мы возьмем дифференциальное уравнение

$$\frac{dy}{dx} = -\frac{x}{y},$$

то увидим, что оно относит каждой отличной от O точке $M(x, y)$ направление, которое перпендикулярно $\overrightarrow{OM}$. Интегральные кривые

в этом случае — окружности с центром в точке O, которая снова является особой точкой нашего дифференциального уравнения, но особой точкой совсем другого типа. Это уже не «узел», а так называемый «центр». Существуют и другие типы особых точек (см. том 2, главу V, § 6), некоторые из них изображены на рис. 21, 22. У дифференциального уравнения $\frac{dy}{dx}=\frac{y}{x}$ нет замкнутых интегральных кривых. Напротив, дифференциальное уравнение $\frac{dy}{dx}=-\frac{x}{y}$ имеет только замкнутые интегральные кривые. Возможны интегральные кривые, спирально закручивающиеся вокруг особой точки, которая в этом случае называется фокусом.

Чрезвычайно важным в разнообразных приложениях является случай так называемого предельного цикла — замкнутой интегральной кривой, на которую спиралевидно накручиваются другие интегральные кривые. Возможны и многие другие случаи взаимного расположения интегральных кривых, а также их расположения по отношению к особым точкам. Все проблемы, касающиеся формы и расположения интегральных кривых дифференциального уравнения, а также числá, характера и взаимного расположения его особых точек, относятся к качественной теории дифференциальных уравнений. Как показывает само название, качественная теория дифференциальных уравнений оставляет в стороне непосредственное интегрирование дифференциального уравнения «в конечном виде», равно как и методы приближенного, численного интегрирования. Основным предметом качественной теории является по существу топология поля направлений и системы интегральных кривых данного дифференциального уравнения.

Качественный подход к дифференциальным уравнениям, включающий такие вопросы, как существование замкнутых интегральных кривых, в частности все вопросы, связанные с существованием, числом и возникновением предельных циклов, продиктован в первую очередь проблемами механики, физики и техники. Впервые эти вопросы возникли в связи с исследованиями Пуанкаре по небесной механике и космогонии, которые, как уже упоминалось, и явились поводом для топологических изысканий французского геометра. В нашей стране топологическая проблематика теории дифференциальных уравнений заняла одно из центральных мест в проведенных советскими учеными выдающихся исследованиях, которые относятся к теории колебаний и радиотехнике — я имею в виду школу Л. И. Мандельштама и выросшую из нее школу А. А. Андронова, ставшую одним из основных центров разработки качественной, в значительной степени топологической теории дифференциальных уравнений. Другим центром разработки качественной теории дифференциальных уравнений в значительной мере топологическими методами является школа В. В. Степанова и В. В. Немыцкого в Московском университете. В работах ленинградских, свердловских и казанских математиков по вопросам

качественной теории дифференциальных уравнений также все возрастающую роль играют топологические методы.

Теория дифференциальных уравнений приводит к изучению векторных полей не только на плоскости, но и в многомерных многообразиях; уже простейшие системы многих дифференциальных уравнений интерпретируются геометрически как поля направлений в многомерных эвклидовых пространствах. Привлечение так называемых первых интегралов уравнения означает выделение среди совокупности всех интегральных кривых тех, которые лежат на некотором многообразии, определяемом данными первыми интегралами. Всякая динамическая система (в классическом смысле этого слова) приводит, вообще говоря, к многомерному многообразию ее возможных состояний (см. § 3) и к системе дифференциальных уравнений, интегральные кривые которой, заполняя данное фазовое пространство, представляют собой возможные движения данной системы. Каждое отдельное из этих движений определяется теми или иными начальными условиями. Следовательно, основным объектом изучения и в этом случае является поле направлений и система траекторий на данном многообразии. Многочисленные приложения, появившиеся особенно в последние годы, делают понятным развитие качественной теории дифференциальных уравнений в ее самом широком аспекте, а следовательно, и развитие топологии как основы этой теории. Именно этой механической и физической, а также астрономической тематике и обязана современная топология своим быстрым ростом, составляющим значительную часть общего развития математики в течение истекшего полустолетия. Читателю, желающему ознакомиться с топологической проблематикой в теории дифференциальных уравнений в ее конкретном физическом и техническом аспекте, можно порекомендовать известную книгу А. А. Андронова и С. Э. Хайкина «Теория колебаний».

В качестве примера конкретной задачи, решенной в теории векторных полей на многомерных многообразиях, рассмотрим вопрос об алгебраическом числе особых точек такого векторного поля.

Пусть нам дано какое-нибудь гладкое многообразие. Для простоты будем представлять себе гладкую замкнутую поверхность. Предположим, что в каждой точке этого многообразия задан касательный к нему вектор, и по длине и по направлению непрерывно зависящий от точки. Особые точки такого векторного поля суть те точки многообразия, в которых отнесенный к ним вектор равен нулю, т. е. не имеет определенного направления. Мы предположим, что каждая из этих точек изолированная. В случае замкнутого многообразия это означает, что имеется лишь конечное число особых точек. (Иначе часть этих точек сгущалась бы около некоторой предельной точки, которая по непрерывности поля также была бы особой точкой, не будучи изолированной.)

Для изолированной особой точки можно определить ее индекс — понятие, в известном смысле аналогичное понятию кратности корня алгеб-

раического уравнения. Для определения индекса окружаем данную особую точку какой-либо «изолирующей ее» замкнутой линией C (на плоскости — просто окружностью), т. е. такой линией, которая сама не проходит ни через одну особую точку, а внутри содержит лишь одну, рассматриваемую нами, особую точку. Во всех точках кривой C направление поля однозначно определено. Для простоты будем считать, что включающая кривую C окрестность нашей особой точки плоская (в общем случае можно отобразить на плоскость интересующую нас окрестность вместе с заданным на ней полем). При обходе кривой в положительном направлении угол, который направление поля образует с каким-либо фиксированным направлением, возвращается к своему начальному значению, возрастая по дороге на некоторое слагаемое вида $2k\pi$, где k — какое-то

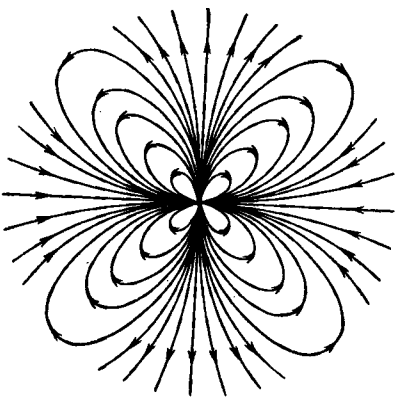


Рис. 21.

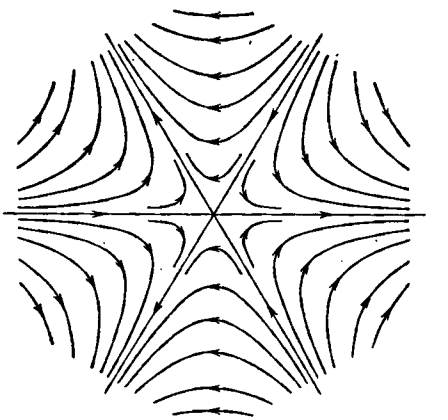


Рис. 22.

целое, вполне определенное число. Это число и называется *индексом особой точки*, а также *вращением* поля вдоль кривой C. Заметим, что оно не зависит от специального выбора замкнутой линии, изолирующей данную особую точку. На рис. 21, 22 изображены особые точки соответственно с индексами $+3$ и -2. Аналогичным, но более сложным образом определяется индекс и для особой точки поля векторов (направлений), определенного на n-мерном многообразии при $n>2$. Оказывается, что имеет место следующая замечательная теорема, доказанная немецким математиком Хопфом в 1926 г.: если на данном многообразии определено непрерывное векторное поле, имеющее лишь конечное число особых точек, то сумма индексов, или, как говорят, алгебраическое число этих особых точек, не зависит от поля и всегда равна эйлеровой характеристике многообразия.

Из теоремы Хопфа следует, что векторные поля без особенностей возможны лишь на таких многообразиях, эйлерова характеристика которых равна нулю; на последних же многообразиях, как оказывается, всегда можно построить векторное поле без особенностей. Таким образом, из всех замкнутых поверхностей только на торе и так называемом

одностороннем торе (поверхности Клейна) можно построить векторное поле, не имеющее особенностей.

С теорией векторных полей тесно связана теория непрерывных отображений многообразий в себя и особенно ее результаты, касающиеся существования неподвижных точек при таких отображениях. Точка x называется неподвижной точкой данного отображения f, если ее образ при этом отображении совпадает с самой точкой, т. е. если

$$f(x)=x.$$

Чтобы пояснить характер этой связи, рассмотрим простейший случай — случай непрерывного отображения f круга K в себя. Соединяя каждую точку x круга K с ее образом $f(x)$, получим вектор $u_x=\overrightarrow{x, f(x)}$. Этот вектор обращается в нуль тогда и только тогда, когда $f(x)=x$, т. е. когда x — неподвижная точка данного отображения. Докажем, что такая точка действительно существует. Для этого предположим противное и определим вращение нашего векторного поля вдоль окружности C круга K.

При непрерывном видоизменении нашего поля его вращение вдоль окружности C, очевидно, может меняться также лишь непрерывно. Являясь при этом целым числом, оно должно оставаться постоянным. Отсюда уже следует, что вращение поля вдоль окружности C равно 1. В самом деле, так как любая точка, принадлежащая кругу K, отображается в этот же круг, то для точки x, лежащей на окружности C, вектор u_x (согласно нашему предположению отличный от нуля) направлен внутрь круга и потому образует острый угол с радиусом Ox, который мы рассматриваем как вектор, направленный к центру O.

Подвергнем непрерывному видоизменению направления всех векторов u_x для точек x, лежащих на окружности C. Видоизменение это будет заключаться в том, что мы повернем все эти векторы на соответствующие острые углы так, чтобы они оказались направленными в центр O. Как было только что сказано, при этом вращение поля вдоль окружности C не изменится. Но в результате такого преобразования наше исходное поле перейдет на C в поле радиальных векторов, которое, очевидно, имеет вращение 1. Итак, наше первоначальное поле также имело вдоль окружности C вращение 1.

В силу непрерывности исходного векторного поля его вращение вдоль двух окружностей с одним и тем же центром O и мало отличающимися по длине радиусами имеет одно и то же значение [1]. Поэтому вращение поля вдоль всех окружностей с центром O, лежащих в круге K, имеет одно и то же значение, а именно 1. Но так как по предположению вектор u_x определен и отличен от нуля во всех точках круга, в том числе и в его

[1] Вследствие предположенного отсутствия у отображения f неподвижных точек рассматриваемое поле всюду определено и отлично от нуля, это и позволяет говорить о его вращении вдоль любой кривой в круге K.

центре O, вращение поля вдоль окружности достаточно малого радиуса с центром в O непременно равно нулю. Мы пришли к противоречию и доказали, таким образом, что при непрерывном отображении круга в себя всегда имеется хотя бы одна неподвижная точка. Эта теорема является частным случаем весьма важной теоремы Брауэра, утверждающей, что при всяком непрерывном отображении в себя n-мерного шара имеется хотя бы одна неподвижная точка.

В настоящее время вопрос о существовании неподвижных точек при отображении тех или иных типов детально изучен и составляет существенную часть топологии многообразий.

§ 6. РАЗВИТИЕ ТОПОЛОГИИ

Топология замкнутых поверхностей — единственная область топологии, которая была более или менее разработана уже к концу прошлого столетия. Построение этой теории было связано с развитием в течение XIX в. теории функций комплексного переменного. Эта последняя, составляя одно из значительнейших явлений в истории математики прошлого века, строилась несколькими различными методами. Одним из наиболее плодотворных в смысле понимания существа изучаемых явлений оказался геометрический метод Римана. Метод Римана, с большой убедительностью показавший, что в общей теории функций комплексного переменного невозможно ограничиться одними лишь однозначными функциями, привел к построению так называемых римановых поверхностей. Эти поверхности в простейшем случае алгебраических функций комплексного переменного всегда оказываются замкнутыми ориентируемыми поверхностями. Изучение их топологических свойств в известном смысле эквивалентно изучению данной алгебраической функции. Дальнейшее развитие идей Римана было произведено Пуанкаре, Клейном и их последователями и привело к установлению неожиданных и глубоких связей между теорией функций, топологией замкнутых поверхностей и неэвклидовой геометрией, а именно — теорией группы движений на плоскости Лобачевского [1]. Таким образом, впервые топология оказалась органически включенной в целый сгусток принципиально значительных проблем, относящихся к весьма различным областям математики.

При дальнейшем развитии этой проблематики оказалось, что одной топологии поверхностей недостаточно, что необходимо решение и определенных задач n-мерной топологии. Первой из них была задача о топологической инвариантности числа измерений пространства. Задача эта заключается в том, чтобы доказать невозможность топологически отобразить n-мерное эвклидово пространство на m-мерное при $n \neq m$. Эта

[1] См. по этому поводу книгу А. И. Маркушевича «Теория аналитических функций». Гостехиздат, 1950.

трудная задача была решена в 1911 г. Брауэром[1]. В связи с ее решением были открыты новые топологические методы, которые привели к быстрому построению начал теории непрерывных отображений многомерных многообразий и теории векторных полей на них. Во всех этих исследованиях оказались необходимыми и первые основные понятия так называемой теоретико-множественной топологии, возникшей на почве общей теории множеств, построенной Кантором в последней четверти прошлого века.

В теоретико-множественной топологии сам объект исследования, т. е. класс рассматриваемых геометрических фигур, чрезвычайно расширился и охватил если не все вообще множества, лежащие в эвклидовых пространствах, то по крайней мере все замкнутые множества. В быстром развитии нового теоретико-множественного направления топологии приняли участие ученые разных стран, причем особо следует отметить польскую топологическую школу.

Существенно новое направление развитие теоретико-множественной топологии получило в работах советских топологов, особенно в построенной выдающимся, безвременно погибшим советским математиком П. С. Урысоном (1898—1924) общей теории размерности, которая заложила основы классификации самых общих точечных множеств по основному признаку — числу измерений. Эта классификация оказалась чрезвычайно плодотворной и повлекла за собой совершенно новые точки зрения в изучении наиболее общих геометрических форм[2]. Идеи Урысона, развитые в его теории размерности, послужили той почвой, на которой возникли замечательные работы Л. А. Люстерника (совместно с Л. Г. Шнирельманом) по вариационному исчислению.

В этих работах наряду с другими результатами было дано исчерпывающее положительное решение знаменитой проблемы Пуанкаре о существовании трех замкнутых годезических линий без кратных точек на всякой поверхности, гомеоморфной сфере.

С другой стороны, на почве теории размерности произошло перенесение П. С. Александровым алгебраических методов комбинаторной топологии в область теории множеств, что в свою очередь повело к новым

[1] В сущности, для развития теории функций комплексного переменного необходимо было решить даже еще более трудную задачу, а именно доказать, что лежащий в n-мерном пространстве топологический образ n-мерной области в свою очередь всегда есть область. Эта задача также была решена Брауэром.

[2] Индуктивное определение размерности множеств, предложенное Урысоном, можно рассматривать как наиболее полное развитие идей Лобачевского о сечении как основной геометрической операции. В самом грубом приближении оно сводится к следующему. Множество нульмерно, если оно может быть представлено в виде суммы сколь угодно малых частей, не находящихся между собой попарно в соприкосновении. Множество n-мерно, если (n — 1)-мерными подмножествами его можно «рассечь» на сколь угодно малые, попарно не соприкасающиеся части и если этого нельзя достигнуть, рассекая его посредством множеств размерности меньше чем n — 1. (Точное определение соприкосновения, как оно понимается в современной топологии, будет дано в § 7.)

направлениям топологических исследований, в которых математики СССР, вплоть до самых молодых, прочно удерживают первое место [1].

Что касается собственно комбинаторной топологии, то после работ Пуанкаре и Брауэра, примерно около 1915 г., начинается цикл исследований американских топологов—Веблена, Биркгофа, Александера, Лефшеца. Ими были достигнуты очень значительные результаты. Так, Александер доказал топологическую инвариантность чисел Бетти, а также свою основную теорему двойственности, послужившую отправной точкой дальнейших исследований Л. С. Понтрягина; Лефшец дал известную формулу об алгебраическом числе неподвижных точек при любых непрерывных отображениях многообразий и тем заложил основы общей алгебраической теории непрерывных отображений, развитой далее Хопфом; Биркгофу наука обязана существенным продвижением теории динамических систем в ее чисто топологическом и в метрическом аспекте и т. д. Дальнейшее очень глубокое развитие топология многообразий и их непрерывных отображений получила в работах Хопфа, который наряду со многими другими результатами доказал существование бесконечного числа непрерывных отображений трехмерной сферы на двумерную, существенно различных между собой в том смысле, что никакие два из этих отображений не могут быть непрерывным видоизменением переведены друг в друга. Хопф, таким образом, становится основателем нового направления — так называемой гомотопической топологии. В настоящее время в гомотопической топологии, как и вообще во всей комбинаторной топологии, произошел новый большой сдвиг, вызванный работами новой французской топологической школы (Лерэ, Серр и др.).

Фундаментальные исследования Урысона были, как уже упоминалось, началом деятельности советских математиков в области топологии. Эти исследования относились к теоретико-множественной топологии, но уже с конца двадцатых годов советские топологи включают в круг своих интересов и комбинаторную топологию. Это включение произошло очень самобытным образом — посредством приложения комбинаторных методов к изучению замкнутых множеств, т. е. объектов очень общей природы. На этой почве произошло одно из значительнейших геометрических открытий текущего столетия — формулировка и доказательство Л. С. Понтрягиным его общего закона двойственности, устанавливающего глубокие и в известном направлении исчерпывающие связи между топологическим строением данного замкнутого множества, лежащего в n-мерном эвклидовом пространстве, и дополнительной к нему части пространства. В связи со своим законом двойственности Л. С. Понтрягин построил общую

[1] Здесь следует указать на так называемую гомологическую теорию размерности П. С. Александрова, на относящиеся к ней замечательные построения Л. С. Понтрягина и на дальнейшее развитие гомологической теории размерности в работах М. Ф. Бокштейна, В. Г. Болтянского и особенно К. А. Ситникова. О законе двойственности Л. С. Понтрягина еще будет идти речь.

теорию характеров коммутативных групп, что привело его и к дальнейшим исследованиям в области общих топологических и классических непрерывных групп Ли — области, которая совершенно преобразована работами Л. С. Понтрягина. В дальнейшем Л. С. Понтрягин и его ученики произвели ряд выдающихся исследований по топологии многообразий и их непрерывных отображений (В. Г. Болтянский, М. М. Постников и др.). В этих исследованиях нашел свое применение новый метод — так называемых ∇-гомологий, введенный в комбинаторную топологию А. Н. Колмогоровым и, независимо от него, Александером. Этот метод, занимающий сейчас первое место во всей гомотопической топологии, позволил в совершенно различных направлениях продолжить теорию двойственности Л. С. Понтрягина, что повело к теоремам двойственности А. Н. Колмогорова (и Александера), П. С. Александрова и К. А. Ситникова, принадлежащим к значительным результатам современной топологии. Этот же метод нашел важные приложения и в новейших работах Л. А. Люстерника по вариационному исчислению.

§ 7. МЕТРИЧЕСКИЕ И ТОПОЛОГИЧЕСКИЕ ПРОСТРАНСТВА

В начале нашего изложения мы говорили о прикосновении (различных частей данной фигуры) как об основном топологическом понятии и определили непрерывные преобразования как такие, которые сохраняют это отношение. Однако точного определения этого основного понятия дано не было; сделать это достаточно общим образом можно лишь на основе понятий теории множеств. Это и составляет задачу настоящего параграфа; ее решение завершается введением понятия топологического пространства.

Теория множеств позволила придать понятию геометрической фигуры широту и общность, недоступную так называемой «классической» математике. Объектом геометрического, в частности топологического, исследования становятся теперь любые точечные множества, т. е. любые множества, элементами которых являются точки n-мерного эвклидова пространства. Между точками n-мерного пространства определено *расстояние*: именно, расстояние между точками $A=(x_1, x_2, \ldots, x_n)$ и $B=(y_1, y_2, \ldots, y_n)$ по определению равно неотрицательному числу

$$\rho(A, B)=\sqrt{(x_1-y_1)^2+(x_2-y_2)^2+\ldots+(x_n-y_n)^2}.$$

Понятие расстояния позволяет определить прикосновение сначала между множеством и точкой, затем — между двумя множествами. Мы говорим, что точка A есть точка прикосновения множества M, если в множестве M имеются точки, расстояние которых от точки A меньше всякого наперед заданного положительного числа. Очевидно, любая точка данного множества есть его точка прикосновения, но и точка, не являющаяся точкой данного множества, может быть его точкой прикосновения. Возьмем, например, открытый промежуток $(0, 1)$ на числовой прямой, т. е.

множество всех точек, лежащих между 0 и 1; сами точки 0 и 1 не принадлежат этому промежутку, но являются его точками прикосновения, так как в промежутке (0, 1) имеются точки, сколь угодно близкие к нулю, и точки, сколь угодно близкие к единице. Множество называется *замкнутым*, если оно содержит все свои точки прикосновения. Например, отрезок [0,1] числовой прямой, т. е. множество всех точек x, удовлетворяющих неравенству $0 \leqslant x \leqslant 1$, замкнуто.

Замкнутые множества на плоскости и тем более в пространстве трех и более измерений могут иметь чрезвычайно сложную структуру; именно они и являются по преимуществу предметом изучения теоретико-множественной топологии n-мерного пространства. Мы скажем далее, что два множества P и Q соприкасаются между собой, если хотя бы одно из них содержит точку прикосновения другого. Согласно этому определению, два пересекающихся (т. е. имеющих общие точки) множества всегда соприкасаются. Из предыдущего следует, что два замкнутых множества соприкасаются лишь в том случае, когда они имеют хотя бы одну общую точку; но, например, отрезок [0, 1] и промежуток (1, 2), не имея общих точек, соприкасаются, так как точка 1, принадлежащая отрезку [0, 1], в то же время является точкой прикосновения промежутка (1, 2). Теперь мы можем сказать, что множество R разбивается («рассекается») лежащим в нем множеством S, или что S является «сечением» множества R, если множество $R—S$, состоящее из всех точек множества R, не принадлежащих S, может быть представлено как сумма двух не соприкасающихся между собой множеств.

Таким образом, идеи Лобачевского о соприкосновении и рассечении множеств получают в современной топологии свое точное и в высшей степени общее выражение. Мы уже видели, как на этих идеях основывается данное Урысоном определение размерности любого множества (см. примечание на стр. 206); формулировка этого определения получает теперь совершенно точное содержание. Это же относится и к определению непрерывного отображения или преобразования: отображение f множества X на множество Y называется непрерывным, если при этом отображении сохраняется соприкосновение, т. е. если из того, что некоторая точка A множества X есть точка прикосновения какого-либо подмножества P множества Y, следует, что образ $f(A)$ точки A есть точка прикосновения образа $f(P)$ множества P. Наконец, взаимно однозначное отображение множества X на множество Y называется *топологическим*, если оно само непрерывно и если обратное ему отображение множества Y на множество X также непрерывно. Эти определения подводят точную базу всему сказанному в первых параграфах настоящей статьи.

Однако теоретико-множественная топология не ограничивается той степенью общности, которая достигается рассмотрением в качестве геометрических фигур всех точечных множеств. Оказывается естественным вводить понятие расстояния не только между точками какого-либо эв-

клидова пространства, но и между другими объектами, казалось бы вовсе не относящимися к геометрии.

Рассмотрим, например, множество всех непрерывных функций, определенных, скажем, на отрезке [0, 1]. Определим расстояние $\rho(f, g)$ между двумя функциями f и g как максимум выражения $|f(x)-g(x)|$, когда x пробегает весь отрезок [0, 1]. Это «расстояние» обладает всеми основными свойствами расстояния между двумя точками пространства: $\rho(f, g)$ между двумя функциями f и g равно нулю тогда и только тогда, когда эти функции совпадают, т. е. когда $f(x)=g(x)$ для любой точки x; далее, расстояние, очевидно, симметрично, т. е. $\rho(f, g)=\rho(g, f)$; наконец, оно удовлетворяет так называемой аксиоме треугольника: для любых трех функций f_1, f_2, f_3 имеем $\rho(f_1, f_2)+\rho(f_2, f_3)\geqslant\rho(f_1, f_3)$. Принято говорить, что приведенное определение расстояния превращает наше множество функций в метрическое пространство (обычно обозначаемое через C). Под метрическим пространством вообще понимается множество каких угодно объектов, условно называемых *точками* метрического пространства, если только между любыми двумя точками определено *расстояние*, являющееся неотрицательным числом, удовлетворяющим приведенным только что «аксиомам расстояния».

Если дано какое-нибудь метрическое пространство, то можно говорить о точках прикосновения его подмножеств, а следовательно, и о соприкосновении его подмножеств между собою и вообще о топологических понятиях (замкнутых множествах, непрерывном отображении и дальнейших понятиях, вводимых на основе этих простейших). Таким путем открывается обширное и чрезвычайно плодотворное поле применения топологических и вообще геометрических идей к таким кругам математических объектов, в применении к которым, казалось бы, совершенно невозможно говорить ни о какой геометрии. Приведем поясняющий пример.

Возьмем снова дифференциальное уравнение (2)

$$\frac{dy}{dx}=F(x, y).$$

Если $y=\varphi(x)$ при $0\leqslant x\leqslant 1$, есть решение этого уравнения, принимающее, положим, для значения $x=0$ значение $y=0$, то функция $\varphi(x)$, очевидно, удовлетворяет интегральному уравнению

$$\varphi(x)=\int_0^x F(x, \varphi(x))\,dx. \tag{3}$$

Рассмотрим теперь интеграл $G(f)=\int_0^x F(x, f(x))\,dx$, где $0\leqslant x\leqslant 1$ и $f(x)$—какая-нибудь непрерывная функция, определенная на отрезке [0, 1].

Этот интеграл есть некоторая непрерывная функция $g(x)$, также определенная на отрезке $[0, 1]$. Так, выражение $G(f)=\int_0^x F(x, f(x))\,dx$ ставит в соответствие каждой функции f функцию $g=G(f)$; другими словами, мы имеем отображение G, как легко видеть — непрерывное, метрического пространства C в себя. Чем же характеризуется при этом функция $\varphi(x)$ (их может быть несколько), являющаяся решением уравнения (2) или эквивалентного ему уравнения (3)? Тем, очевидно, что при нашем отображении она переходит сама в себя, т. е. является неподвижной точкой нашего отображения G. Оказывается, такая неподвижная точка у отображения G действительно существует — это вытекает из весьма общей теоремы о неподвижных точках непрерывных отображений метрических пространств, доказанной в 1926 г. П. С. Александровым и В. В. Немыцким. В настоящее время рассмотрение различных метрических пространств, точками которых являются те или иные функции (такие пространства называются функциональными), является постоянно применяемым аппаратом анализа, а изучение функциональных пространств посредством отчасти топологических, а главным образом, в широком смысле слова, алгебраических методов, составляет содержание функционального анализа (см. главу XIX).

Функциональный анализ, как уже упоминалось в вводной главе, занимает в современной математической науке чрезвычайно видное место ввиду многообразия своих связей со всевозможными другими частями математики и по своему значению для естествознания, в первую очередь для теоретической физики. Исследование топологических свойств функциональных пространств тесно связано с вариационным исчислением и теорией уравнений с частными производными (исследования Л. А. Люстерника, Морса, Лерэ, Шаудера, М. А. Красносельского и др.). Вопросы существования неподвижных точек при непрерывных отображениях функциональных пространств занимают в этих исследованиях значительное место.

Топология функциональных и вообще метрических пространств — всё еще не последнее слово общности в современных топологических теориях. Дело в том, что в метрических пространствах основное топологическое понятие прикосновения вводится на основе расстояния между точками, которое само по себе не является топологическим понятием. Поэтому возникает задача непосредственного, аксиоматического определения прикосновения. Так мы приходим к понятию топологического пространства — самому общему понятию современной топологии.

Топологическое пространство есть множество объектов любой природы (называемых точками топологического пространства), в котором для каждого подмножества тем или иным способом заданы его точки

прикосновения. При этом требуется соблюдение некоторых естественных условий, называемых аксиомами топологического пространства (например, каждая точка данного множества является его точкой прикосновения, точка прикосновения суммы двух множеств является точкой прикосновения по крайней мере одного из слагаемых и т. п.). Теория топологических пространств — в настоящее время глубоко разработанная область математики. В ее развитии приняли ведущее участие советские математики П. С. Урысон, П. С. Александров, А. Н. Тихонов и др. Из новейших результатов в теории топологических пространств, имеющих принципиальное значение, следует отметить найденное молодым математиком Ю. М. Смирновым необходимое и достаточное условие метризуемости топологического пространства, т. е. условие, при котором между точками пространства можно определить расстояние таким образом, чтобы имеющаяся «топология» пространства могла рассматриваться как порожденная этим понятием расстояния. Иными словами, чтобы точки прикосновения всевозможных множеств в полученном метрическом пространстве были теми же самыми, что и определенные с самого начала в данном топологическом пространстве.

ЛИТЕРАТУРА

Александров П. С. Комбинаторная топология. Гостехиздат, 1947.

Обширное систематическое руководство, предъявляющее довольно большие требования к читателю.

Александров П. С. Теоремы двойственности в комбинаторной топологии. Юб. сборник, посвящ. 30-летию Великой Октябрьской социалистической революции, т. 1, Изд-во АН СССР, 1947.

Александров П. С. Введение в общую теорию множеств и функций. Гостехиздат, 1948.

Эта книга является и введением в теоретико-множественную топологию (метрических и топологических пространств).

Александров П. С. Топологические теоремы двойственности, ч. 1, Замкнутые множества (Труды Математического института им. В. А. Стеклова АН СССР, т. 48, 1955).

Может служить первой книгой для чтения по комбинаторной топологии, доводящей читателя до современных исследований в одной из важных областей топологии (теории двойственности).

Александров П. С. и Ефремович В. А. Очерк основных понятий топологии. М.—Л., 1936.

Книга представляет собою введение в комбинаторную топологию с довольно подробным изложением топологии поверхностей.

Понтрягин Л. С. Основы комбинаторной топологии. Гостехиздат, 1947.

Очень сжатое, но безукоризненно полное изложение основных предложений n-мерной комбинаторной топологии.

Глава XIX

ФУНКЦИОНАЛЬНЫЙ АНАЛИЗ

Возникновение и развитие в XX в. функционального анализа обусловлено в основном двумя причинами. С одной стороны, появилась необходимость осмыслить с единой точки зрения богатейший фактический материал, накопленный в различных, часто мало связанных между собой, разделах математики в течение XIX в. При этом основные понятия функционального анализа возникали и выкристаллизовывались с разных сторон и по разным поводам. Многие из основных понятий функционального анализа появились естественным образом в процессе развития вариационного исчисления, в задачах о колебаниях (при переходе от колебаний систем с конечным числом степеней свободы к колебаниям сплошных сред), в теории интегральных уравнений, в теории дифференциальных уравнений, обыкновенных и с частными производными (в краевых задачах, в задачах о собственных значениях и т. д.), при развитии теории функций действительного переменного, в операционном исчислении, при рассмотрении задач теории приближения функций и др. Функциональный анализ позволил понять с единой точки зрения многие имеющиеся в этих областях результаты и часто способствовал получению новых. Подготовленные при этом понятия и аппарат были использованы в последние десятилетия в новом разделе теоретической физики — в квантовой механике.

С другой стороны, изучение математических проблем, связанных с квантовой механикой, явилось переломным моментом в ходе дальнейшего развития самого функционального анализа и сформировало и формирует в настоящее время основные направления его развития.

Функциональный анализ еще далек от своего завершения. Однако несомненно, что и в дальнейшем его развитии задачи и потребности современной физики будут иметь для него такое же значение, какое классическая механика имела для возникновения и развития в XVIII в. дифференциального и интегрального исчисления.

Мы далеки от мысли охватить в этой главе все, хотя бы основные, вопросы функционального анализа. Многие важные разделы остались за пределами этой статьи. Мы хотели тем не менее, ограни-

чившись несколькими избранными вопросами, познакомить читателя с некоторыми основными понятиями функционального анализа и проиллюстрировать по мере возможности те связи, о которых мы здесь упомянули. Эти вопросы по существу разобраны в начале XX в. в основном в классических работах Гильберта — одного из основоположников функционального анализа. С тех пор функциональный анализ очень сильно развился и широко применяется почти во всех разделах математики: в дифференциальных уравнениях в частных производных, в теории вероятностей, в квантовой механике, в квантовой теории полей и т. д. К сожалению, это дальнейшее развитие функционального анализа здесь не отражено. Для того, чтобы описать это дальнейшее развитие нужно было бы написать отдельную большую статью и мы ограничиваемся поэтому одним из наиболее старых вопросов — теорией собственных функций.

§ 1. n-МЕРНОЕ ПРОСТРАНСТВО

В дальнейшем мы будем пользоваться основными понятиями, относящимися к n-мерному пространству. Хотя в главах о линейной алгебре и абстрактных пространствах эти понятия и вводились, мы считаем нелишним повторить их здесь коротко в том виде, в котором они далее встретятся. При чтении этого параграфа читателю достаточно знания основ аналитической геометрии.

Мы знаем, что точка в аналитической геометрии трехмерного пространства задается тройкой чисел (f_1, f_2, f_3) — координат точки. Расстояние этой точки до начала координат равно $\sqrt{f_1^2+f_2^2+f_3^2}$. Если считать точку концом вектора, ведущего в эту точку из начала координат, то длина вектора также равна $\sqrt{f_1^2+f_2^2+f_3^2}$. Косинус угла между ненулевыми векторами, ведущими из начала координат в две различные точки $A(f_1, f_2, f_3)$ и $B(g_1, g_2, g_3)$, определяется формулой

$$\cos\varphi = \frac{f_1g_1+f_2g_2+f_3g_3}{\sqrt{f_1^2+f_2^2+f_3^2}\sqrt{g_1^2+g_2^2+g_3^2}}.$$

Из тригонометрии мы знаем, что $|\cos\varphi| \leqslant 1$. Поэтому имеет место неравенство

$$\frac{|f_1g_1+f_2g_2+f_3g_3|}{\sqrt{f_1^2+f_2^2+f_3^2}\sqrt{g_1^2+g_2^2+g_3^2}} \leqslant 1,$$

и, значит, всегда

$$(f_1g_1+f_2g_2+f_3g_3)^2 \leqslant (f_1^2+f_2^2+f_3^2)(g_1^2+g_2^2+g_3^2). \qquad (1)$$

Это последнее неравенство носит алгебраический характер и справедливо для любых шести чисел (f_1, f_2, f_3) и (g_1, g_2, g_3), так как любые шесть чисел могут служить координатами двух точек пространства.

В то же время неравенство (1) получено из чисто геометрических соображений и тесно связано с геометрией, что позволяет сделать его наглядным.

При аналитической формулировке ряда геометрических соотношений часто оказывается, что соответствующие факты остаются справедливыми при замене тройки чисел на n чисел. Например, выведенное нами неравенство (1) может быть обобщено на $2n$ чисел $(f_1, f_2, \ldots, f_n)$ и $(g_1, g_2, \ldots, g_n)$. Это значит, что для любых $2n$ чисел $(f_1, f_2, \ldots, f_n)$ и $(g_1, g_2, \ldots, g_n)$ справедливо неравенство, аналогичное (1), а именно:

$$(f_1g_1+f_2g_2+\ldots+f_ng_n)^2 \leqslant (f_1^2+f_2^2+\ldots+f_n^2)(g_1^2+g_2^2+\ldots+g_n^2). \quad (1')$$

Это неравенство, частным случаем которого является неравенство (1), может быть доказано чисто аналитически[1]. Подобным образом обобщаются на n чисел многие другие соотношения между тройками чисел, полученные из аналитической геометрии. Такая связь соотношений между числами (количественных соотношений), примером которых является указанное выше неравенство, с геометрией становится особенно выпуклой, когда вводится понятие n-мерного пространства. Это понятие было введено в главе XVI. Мы здесь вкратце его повторим.

Совокупность n чисел $(f_1, f_2, \ldots, f_n)$ называют точкой или *вектором* n-мерного пространства (мы чаще будем употреблять последнее название). Вектор $(f_1, f_2, \ldots, f_n)$ будем в дальнейшем коротко обозначать одной буквой f.

Подобно тому, как в трехмерном пространстве при сложении векторов складываются их компоненты, сумма векторов

$$f=\{f_1, f_2, \ldots, f_n\} \text{ и } g=\{g_1, g_2, \ldots, g_n\}$$

определяется как вектор $\{f_1+g_1, f_2+g_2, \ldots, f_n+g_n\}$ и обозначается $f+g$.

Произведение вектора $f=\{f_1, f_2, \ldots, f_n\}$ на число λ есть вектор $\lambda f=\{\lambda f_1, \lambda f_2, \ldots, \lambda f_n\}$.

Длина вектора $f=\{f_1, f_2, \ldots, f_n\}$, аналогично длине вектора в трехмерном пространстве, определяется как $\sqrt{f_1^2+f_2^2+\ldots+f_n^2}$.

Угол φ между двумя векторами $f=\{f_1, f_2, \ldots, f_n\}$ и $g=\{g_1, g_2, \ldots, g_n\}$ в n-мерном пространстве задается своим косинусом, также совершенно аналогично тому, как определяется угол между векторами в трехмерном пространстве. Именно, он определяется формулой

$$\cos\varphi=\frac{f_1g_1+f_2g_2+\ldots+f_ng_n}{\sqrt{f_1^2+f_2^2+\ldots+f_n^2}\sqrt{g_1^2+g_2^2+\ldots+g_n^2}}.^2 \quad (2)$$

Скалярным произведением двух векторов называется произведение длин этих векторов на косинус угла между ними. Таким образом, если

[1] См. стр. 56—57.

[2] То, что $|\cos\varphi| \leqslant 1$, следует из неравенства (1').

$f=\{f_1, f_2, \ldots, f_n\}$ и $g=\{g_1, g_2, \ldots, g_n\}$, то, так как длины векторов равны соответственно $\sqrt{f_1^2+f_2^2+\ldots+f_n^2}$ и $\sqrt{g_1^2+g_2^2+\ldots+g_n^2}$, их скалярное произведение, которое обозначается через (f, g), задается формулой

$$(f, g)=f_1g_1+f_2g_2+\ldots+f_ng_n. \qquad (3)$$

В частности, условием ортогональности (перпендикулярности) двух векторов является равенство $\cos\varphi=0$, т. е. $(f, g)=0$.

При помощи формулы (3) читатель может убедиться, что скалярное произведение в n-мерном пространстве обладает следующими свойствами:

1. $(f, g)=(g, f)$.
2. $(\lambda f, g)=\lambda(f, g)$.

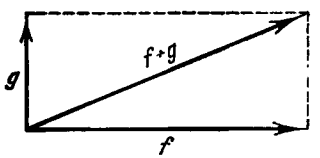


Рис. 1.

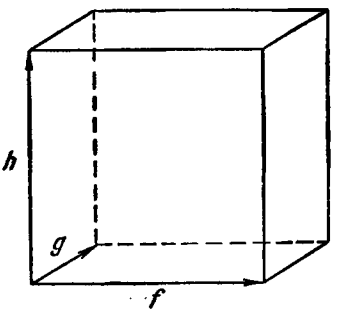


Рис. 2.

3. $(f, g_1+g_2)=(f, g_1)+(f, g_2)$.
4. $(f, f)\geqslant 0$, причем знак равенства имеет место, только при $f=0$, т. е. когда $f_1=f_2=\ldots=f_n=0$.

Скалярное произведение вектора f самого на себя (f, f) равно квадрату длины вектора f.

Скалярное произведение оказывается очень удобным средством при изучении n-мерных пространств. Мы не будем здесь изучать геометрию n-мерного пространства, а ограничимся лишь одним примером.

В качестве такого примера рассмотрим в n-мерном пространстве теорему Пифагора: квадрат гипотенузы равен сумме квадратов катетов. Для этого выберем такое доказательство этой теоремы на плоскости, которое легко переносится на случай n-мерного пространства.

Пусть f и g — два перпендикулярных вектора на плоскости. Рассмотрим прямоугольный треугольник, построенный на векторах f и g (рис. 1). Гипотенуза этого треугольника равна по длине вектору $f+g$. Запишем векторно теорему Пифагора в наших обозначениях. Так как квадрат длины вектора равен скалярному произведению вектора на себя, то в терминах скалярных произведений теорема Пифагора запишется так:

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

Доказательство непосредственно следует из свойств скалярного произведения. Действительно,

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

а два средних слагаемых равны нулю в силу ортогональности векторов f и g.

В приведенном выше доказательстве мы пользовались лишь определением длины вектора, перпендикулярности векторов и свойствами скалярного произведения. Поэтому в доказательстве ничего не изменится, если мы предположим, что f и g — два ортогональных вектора n-мерного пространства. Тем самым теорема Пифагора доказана для прямоугольного треугольника в n-мерном пространстве.

Если задано три попарно ортогональных вектора в n-мерном пространстве f, g и h, то сумма этих векторов $f+g+h$ является диагональю прямоугольного параллелепипеда, построенного на этих векторах (рис. 2), и имеет место равенство

$$(f+g+h,\ f+g+h)=(f,\ f)+(g,\ g)+(h,\ h),$$

которое означает, что квадрат длины диагонали параллелепипеда равен сумме квадратов длин его ребер. Доказательство этого утверждения, вполне аналогичное приведенному выше доказательству теоремы Пифагора, мы предоставляем читателю. Точно так же, если в n-мерном пространстве имеется k попарно ортогональных векторов $f^1, f^2, \ldots, f^k$, то столь же просто доказываемое равенство

$$(f^1+f^2+\ldots+f^k,\ f^1+f^2+\ldots+f^k)=(f^1,\ f^1)+(f^2,\ f^2)+\ldots+(f^k,\ f^k) \quad (4)$$

означает, что квадрат длины диагонали «k-мерного параллелепипеда» в n-мерном пространстве также равен сумме квадратов длин его ребер.

§ 2. ГИЛЬБЕРТОВО ПРОСТРАНСТВО (БЕСКОНЕЧНОМЕРНОЕ ПРОСТРАНСТВО)

Связь с n-мерным пространством. Введение понятия n-мерного пространства оказалось полезным при изучении ряда вопросов математики и физики. В свою очередь это понятие дало толчок дальнейшему развитию понятия пространства и его приложению к различным областям математики. В развитии линейной алгебры и геометрии n-мерных пространств большую роль сыграли задачи о малых колебаниях упругих систем.

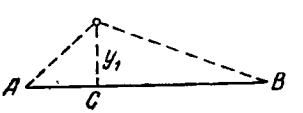


Рис. 3.

Рассмотрим следующий классический пример такой задачи (рис. 3). Пусть AB — гибкая нить, натянутая между точками A и B. Представим себе, что в некоторой точке C нити прикреплен грузик. Если вывести его из положения равновесия, он начнет колебаться с некоторой частотой ω, которую можно вычислить, зная силу натяжения нити,

массу m и положение груза. Положение системы в каждый момент времени задается при этом одним числом, а именно отклонением y_1 массы m от положения равновесия нити.

Расположим теперь на нити AB n грузиков в точках $C_1, C_2, \ldots, C_n$. Нить при этом будем считать невесомой. Это значит, что ее масса настолько мала, что по сравнению с массами грузиков мы можем пренебречь ею. Положение такой системы задается n числами $y_1, y_2, \ldots, y_n$, равными отклонениям грузиков от положения равновесия. Совокупность чисел $y_1, y_2, \ldots, y_n$ можно (и это оказывается во многих отношениях полезным) считать вектором $(y_1, y_2, \ldots, y_n)$ n-мерного пространства.

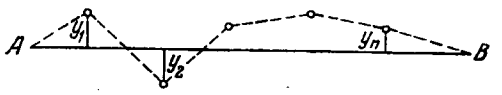


Рис. 4.

Само исследование малых колебаний, проведенное в таком изложении, оказывается тесно связанным с основными фактами геометрии n-мерных пространств. Укажем, например, что определение частот колебаний такой системы может быть сведено к задаче о нахождении осей некоторого эллипсоида в n-мерном пространстве.

Рассмотрим теперь задачу о малых колебаниях натянутой между точками A и B струны. При этом мы имеем в виду идеализированную струну, т. е. гибкую нить, обладающую конечной массой, непрерывно распределенной вдоль нити. В частности, под однородной струной понимается струна, плотность которой постоянна.

Так как масса распределена вдоль струны непрерывно, то положение струны уже нельзя задать конечным числом чисел $y_1, y_2, \ldots, y_n$, а нужно задать отклонение $y(x)$ каждой точки x струны. Таким образом, положение струны в каждый момент времени задается некоторой функцией $y(x)$.

Положение нити с n грузами, прикрепленными в точках с абсциссами $x_1, x_2, \ldots, x_n$, изображается графически ломаной линией с n звеньями (рис. 4). Если число грузов увеличится, то соответствующим образом увеличится число звеньев ломаной линии. Если число грузов неограниченно возрастает, а расстояние между соседними грузами стремится

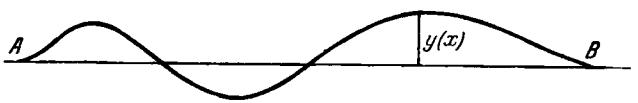


Рис. 5.

к нулю, то в пределе мы получим непрерывное распределение масс вдоль нити, т. е. идеализированную струну. Ломаная линия, изображающая положение нити с грузами, перейдет при этом в кривую, изображающую положение струны (рис. 5).

Мы видим, таким образом, что между задачами о колебании нити с грузами и о колебании струны существует тесная связь. В первой задаче положение системы задавалось точкой или вектором n-мерного пространства. Поэтому естественно функцию $f(x)$, задающую положение колеблющейся струны во второй задаче, также рассматривать как вектор или точку некоторого, уже бесконечномерного пространства. К той же идее рассмотрения пространства, точками (векторами) которого являются функции $f(x)$, заданные на некотором интервале, приводит целый ряд аналогичных задач[1].

Рассмотренный выше пример задачи о колебаниях, к которому мы еще вернемся в § 4, подсказывает нам, как следует вводить основные понятия в бесконечномерном пространстве.

Гильбертово пространство. Мы рассмотрим здесь одно из наиболее распространенных и важных для приложений понятий бесконечномерного пространства, а именно понятие гильбертова пространства.

Вектор n-мерного пространства определялся как совокупность n чисел f_i, где i меняется от 1 до n. Аналогично, вектор бесконечномерного пространства определяется как функция $f(x)$, где x меняется от a до b.

Сложение векторов и умножение вектора на число определяется как сложение функций и умножение функции на число.

Длина вектора f в n-мерном пространстве определялась формулой $\sqrt{\sum_{i=1}^{n} f_i^2}$. Так как для функций роль суммы играет интеграл, то длина вектора $f(x)$ гильбертова пространства задается формулой

$$\sqrt{\int_a^b f^2(x)\,dx}. \tag{5}$$

[1] В качестве еще одной такой задачи рассмотрим электрические колебания, возникающие в цепочке связанных между собой электрических контуров (рис. 6).

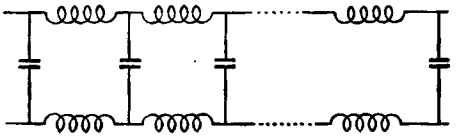


Рис. 6.

Состояние такой цепочки выражается набором n чисел $u_1, u_2, \ldots, u_n$, где u_i — напряжение на конденсаторе i-го контура цепочки. Совокупность n чисел $(u_1, \ldots, u_n)$ является вектором n-мерного пространства.

Представим себе теперь двухпроводную линию, т. е. линию, состоящую из двух проводов, имеющих конечную емкость и индуктивность, распределенные вдоль линии. Электрическое состояние линии выражается некоторой функцией $u(x)$, задающей распределение напряжения вдоль линии. Эта функция является вектором бесконечномерного пространства функций, заданных на интервале (a, b).

Расстояние между точками f и g в n-мерном пространстве определяется как длина вектора $f-g$, т. е. как $\sqrt{\sum_{i=1}^{n}(f_i-g_i)^2}$. Аналогично «расстояние» между элементами $f(t)$ и $g(t)$ в функциональном пространстве равно $\sqrt{\int_a^b [f(t)-g(t)]^2\,dt}$.

Выражение $\int_a^b [f(t)-g(t)]^2\,dt$ называется средним квадратичным уклонением функций $f(t)$ и $g(t)$. Таким образом, за меру расстояния двух элементов гильбертова пространства принимается их среднее квадратичное уклонение.

Перейдем к определению угла между векторами. В n-мерном пространстве угол φ между векторами $f=\{f_i\}$ и $g=\{g_i\}$ определяется формулой

$$\cos\varphi=\frac{\sum_{i=1}^{n} f_i g_i}{\sqrt{\sum_{i=1}^{n} f_i^2}\sqrt{\sum_{i=1}^{n} g_i^2}}.$$

В бесконечномерном пространстве суммы заменяются соответствующими интегралами и угол φ между двумя векторами f и g гильбертова пространства будет определяться аналогичной формулой

$$\cos\varphi=\frac{\int_a^b f(t)\,g(t)\,dt}{\sqrt{\int_a^b f^2(t)\,dt}\sqrt{\int_a^b g^2(t)\,dt}}. \tag{6}$$

Такое выражение можно считать косинусом некоторого угла φ, если стоящая справа дробь по абсолютной величине меньше единицы, т. е. если

$$\left|\int_a^b f(t)\,g(t)\,dt\right|\leqslant\sqrt{\int_a^b f^2(t)\,dt}\sqrt{\int_a^b g^2(t)\,dt}\,. \tag{7}$$

Это неравенство действительно имеет место для любых двух функций $f(t)$ и $g(t)$. Оно играет важную роль в анализе и носит название неравенства Коши — Буняковского. Докажем его.

Пусть $f(x)$ и $g(x)$ — две функции, нетождественно равные нулю, заданные на интервале (a, b). Возьмем произвольные числа λ и μ и составим выражение

$$\int_a^b [\lambda f(x) - \mu g(x)]^2 dx.$$

Так как функция $[\lambda f(x) - \mu g(x)]^2$, стоящая под знаком интеграла, неотрицательна, то имеет место неравенство

$$\int_a^b [\lambda f(x) - \mu g(x)]^2 dx \geqslant 0,$$

т. е.

$$2\lambda\mu \int_a^b f(x) g(x) dx \leqslant \lambda^2 \int_a^b f^2(x) dx + \mu^2 \int_a^b g^2(x) dx.$$

Введем для краткости обозначения

$$\left| \int_a^b f(x) g(x) dx \right| = C, \quad \int_a^b f^2(x) dx = A, \quad \int_a^b g^2(x) dx = B. \tag{8}$$

В этих обозначениях неравенство перепишется следующим образом[1]:

$$2\lambda\mu C \leqslant \lambda^2 A + \mu^2 B. \tag{9}$$

Это неравенство остается справедливым для любых значений λ и μ; в частности, мы можем положить

$$\lambda = \sqrt{\frac{C}{A}}, \quad \mu = \sqrt{\frac{C}{B}}. \tag{10}$$

Подставив эти значения λ и μ в неравенство (9), получаем

$$\frac{C}{\sqrt{AB}} \leqslant 1.$$

Заменяя A, B и C их выражениями по формуле (8), получаем окончательно неравенство Коши — Буняковского.

В геометрии скалярное произведение векторов определяется как произведение их длин на косинус угла между ними. Длины векторов f и g в нашем случае равны

$$\sqrt{\int_a^b f^2(x) dx} \quad \text{и} \quad \sqrt{\int_a^b g^2(x) dx},$$

[1] Мы можем взять в качестве C модуль интеграла за счет произвольности в знаке λ или μ.

а косинус угла между ними определяется формулой

$$\cos\varphi = \frac{\int_a^b f(x)\,g(x)\,dx}{\sqrt{\int_a^b f^2(x)\,dx}\sqrt{\int_a^b g^2(x)\,dx}}.$$

Перемножая эти выражения, мы приходим к следующей формуле для скалярного произведения двух векторов гильбертова пространства:

$$(f,\ g) = \int_a^b f(x)\,g(x)\,dx. \tag{11}$$

Из этой формулы видно, что скалярное произведение вектора f на себя есть квадрат его длины.

Если скалярное произведение ненулевых векторов f и g равно нулю, то это значит, что $\cos\varphi = 0$, т. е. приписанный им нашим определением угол φ равен 90°. Поэтому функции f и g, для которых

$$(f,\ g) = \int_a^b f(x)\,g(x)\,dx = 0,$$

называются ортогональными.

В гильбертовом пространстве, как и в n-мерном, справедлива теорема Пифагора (см. § 1). Пусть $f_1(x)$, $f_2(x)$, ..., $f_N(x)$ представляют собой N попарно ортогональных функций, а

$$f(x) = f_1(x) + f_2(x) + \ldots + f_N(x)$$

их сумму. Тогда квадрат длины f равен сумме квадратов длин f_1, f_2, ..., f_N.

Так как длины векторов в гильбертовом пространстве задаются при помощи интегралов, то теорема Пифагора в этом случае выражается формулой

$$\int_a^b f^2(x)\,dx = \int_a^b f_1^2(x)\,dx + \int_a^b f_2^2(x)\,dx + \ldots + \int_a^b f_N^2(x)\,dx. \tag{12}$$

Доказательство этой теоремы ничем не отличается от изложенного выше (§ 1) доказательства той же теоремы в n-мерном пространстве.

Мы не уточнили до сих пор, какие именно функции следует считать векторами гильбертова пространства. В качестве таких функций следует взять все функции, для которых имеет смысл интеграл $\int_a^b f^2(x)\,dx$.

Казалось бы естественным ограничиться непрерывными функциями, для

которых $\int_a^b f^2(x)\,dx$ заведомо всегда существует. Однако наибольшую законченность и естественность теория гильбертова пространства получает, если интеграл $\int_a^b f^2(x)\,dx$ понимать в обобщенном смысле, а именно в смысле так называемого интеграла Лебега (см. главу XV).

Это расширение понятия интеграла (и соответственно — класса рассматриваемых функций) необходимо для функционального анализа так же, как для обоснования дифференциального и интегрального исчисления необходима строгая теория действительных чисел. Таким образом, созданное в начале XX в. в связи с развитием теории функций действительного переменного обобщение обычного понятия интеграла оказалось весьма существенным для функционального анализа и связанных с ним разделов математики.

§ 3. РАЗЛОЖЕНИЕ ПО ОРТОГОНАЛЬНЫМ СИСТЕМАМ ФУНКЦИЙ

Определение и примеры ортогональных систем функций. Если на плоскости выбрать какие-нибудь два взаимно перпендикулярных вектора e_1 и e_2 единичной длины (рис. 7), то произвольный вектор в той же плоскости можно разложить по направлениям этих двух векторов, т. е. представить его в виде

$$f = a_1 e_1 + a_2 e_2,$$

где a_1 и a_2 — числа, равные проекциям вектора f на направления осей e_1 и e_2. Так как проекция f на ось равна произведению длины f на косинус угла f с осью, то, вспоминая определение скалярного произведения, мы можем написать

$$a_1 = (f,\ e_1),$$
$$a_2 = (f,\ e_2).$$

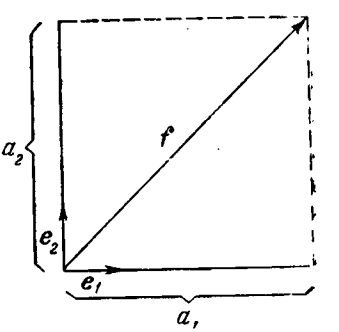


Рис. 7.

Аналогично, если в трехмерном пространстве выбрать какие-нибудь три взаимно перпендикулярных вектора e_1, e_2, e_3 единичной длины, то произвольный вектор f в этом пространстве можно представить в виде

$$f = a_1 e_1 + a_2 e_2 + a_3 e_3,$$

где

$$a_k = (f,\ e_k) \quad (k = 1,\ 2,\ 3).$$

В гильбертовом пространстве также можно рассматривать системы попарно ортогональных векторов этого пространства, т. е. функций $\varphi_1(x)$, $\varphi_2(x)$, ..., $\varphi_n(x)$, ...

Такие системы функций называются ортогональными системами функций и играют большую роль в анализе. Они встречаются в самых различных вопросах математической физики, интегральных уравнений, приближенных вычислений, теории функций действительного переменного и т. п. Упорядочение и объединение понятий, относящихся к таким системам, были одним из стимулов, приведших в начале XX в. к созданию общего понятия гильбертова пространства.

Дадим точные определения. Система функций

$$\varphi_1(x),\ \varphi_2(x),\ \ldots,\ \varphi_n(x),\ \ldots$$

называется *ортогональной*, если любые две функции этой системы ортогональны между собой, т. е. если

$$\int_a^b \varphi_i(x)\,\varphi_k(x)\,dx = 0 \text{ для } i \neq k. \tag{13}$$

В трехмерном пространстве мы требовали, чтобы длины векторов системы равнялись единице. Вспомнив определение длины вектора, мы видим, что в случае гильбертова пространства это требование записывается так:

$$\int_a^b \varphi_k^2(x)\,dx = 1. \tag{14}$$

Система функций, удовлетворяющая требованиям (13) и (14), называется ортогональной и нормированной.

Приведем примеры таких систем функций.

1. На интервале $(-\pi, \pi)$ рассмотрим последовательность функций

$$1,\ \cos x,\ \sin x,\ \cos 2x,\ \sin 2x,\ \ldots,\ \cos nx,\ \sin nx,\ \ldots$$

Каждые две функции из этой последовательности ортогональны между собой. Это проверяется простым вычислением соответствующих интегралов. Квадрат длины вектора в гильбертовом пространстве есть интеграл от квадрата функции. Таким образом, квадраты длин векторов последовательности

$$1,\ \cos x,\ \sin x,\ \cos 2x,\ \sin 2x,\ \ldots,\ \cos nx,\ \sin nx,\ \ldots$$

суть интегралы

$$\int_{-\pi}^{\pi} dx = 2\pi,\quad \int_{-\pi}^{\pi} \cos^2 nx\,dx = \pi,\quad \int_{-\pi}^{\pi} \sin^2 nx\,dx = \pi,$$

т. е. последовательность наших векторов ортогональна, но не нормирована. Длина первого вектора последовательности равна $\sqrt{2\pi}$, а все

остальные имеют длину $\sqrt{\pi}$. Поделив каждый вектор на его длину, мы получим ортогональную и нормированную систему тригонометрических функций

$$\frac{1}{\sqrt{2\pi}}, \frac{\cos x}{\sqrt{\pi}}, \frac{\sin x}{\sqrt{\pi}}, \frac{\cos 2x}{\sqrt{\pi}}, \frac{\sin 2x}{\sqrt{\pi}}, \ldots, \frac{\cos nx}{\sqrt{\pi}}, \frac{\sin nx}{\sqrt{\pi}}, \ldots$$

Эта система является исторически одним из первых и наиболее важных примеров ортогональных систем. Она возникла в работах Эйлера, Д. Бернулли, Даламбера в связи с задачей о колебаниях струны. Ее изучение сыграло существенную роль в развитии всего анализа[1].

Появление ортогональной системы тригонометрических функций в связи с задачей о колебаниях струны не случайно. Каждая задача о малых колебаниях среды приводит к некоторой системе ортогональных функций, описывающих так называемые собственные колебания данной системы (см. § 4). Например, в связи с задачей о колебаниях сферы появляются так называемые сферические функции, в связи с задачей о колебаниях круглой мембраны или цилиндра появляются так называемые цилиндрические функции и т. д.

2. Можно привести пример ортогональной системы функций, каждая функция которой является многочленом. Таким примером является последовательность многочленов Лежандра

$$P_n(x) = \frac{1}{2^n n!} \frac{d^n (x^2-1)^n}{dx^n},$$

т. е. $P_n(x)$ есть (с точностью до постоянного множителя) производная n-го порядка от $(x^2-1)^n$. Выпишем первые несколько многочленов этой последовательности:

$$P_0(x) = 1;$$
$$P_1(x) = x;$$
$$P_2(x) = \frac{1}{2}(3x^2-1);$$
$$P_3(x) = \frac{1}{2}(5x^3-3x).$$

Очевидно, что вообще $P_n(x)$ есть многочлен n-й степени. Мы предоставляем читателю самому убедиться, что эти многочлены представляют собой ортогональную последовательность на интервале $(-1, 1)$.

Общую теорию ортогональных многочленов (так называемые ортогональные многочлены с весом) развил замечательный русский математик П. Л. Чебышев во второй половине XIX в.

Разложение по ортогональным системам функций. Подобно тому как в трехмерном пространстве каждый вектор можно представить

[1] См. главу XII (том 2), § 1.

в виде линейной комбинации трех попарно ортогональных векторов e_1, e_2, e_3 единичной длины

$$f = a_1 e_1 + a_2 e_2 + a_3 e_3,$$

в функциональном пространстве возникает задача о разложении произвольной функции f в ряд по ортогональной и нормированной системе функций, т. е. о представлении функции f в виде

$$f(x) = a_1 \varphi_1(x) + a_2 \varphi_2(x) + \ldots + a_n \varphi_n(x) + \ldots \qquad (15)$$

При этом сходимость ряда (15) к функции f понимается в смысле расстояния между элементами в гильбертовом пространстве. Это значит, что среднее квадратичное уклонение частичной суммы ряда $S_n(t) = \sum_{k=1}^{n} a_k \varphi_k(t)$ от функции $f(t)$ стремится к нулю при $n \to \infty$, т. е.

$$\lim_{n\to\infty} \int_a^b [f(t) - S_n(t)]^2 \, dt = 0. \qquad (16)$$

Такая сходимость называется обычно «сходимостью в среднем».

Разложения по тем или иным системам ортогональных функций часто встречаются в анализе и являются важным методом для решения задач математической физики. Так, например, если ортогональная система есть система тригонометрических функций на интервале $(-\pi, \pi)$

$$1, \cos x, \sin x, \cos 2x, \sin 2x, \ldots, \cos nx, \sin nx, \ldots,$$

то такое разложение есть классическое разложение функции в тригонометрический ряд[1]

$$f(x) = a_0 + a_1 \cos x + b_1 \sin x + a_2 \cos 2x + b_2 \sin 2x + \ldots$$

Предположим, что разложение (15) возможно для любой функции f из гильбертова пространства, и найдем коэффициенты a_n такого разложения. Для этого умножим обе части равенства скалярно на одну и ту же функцию φ_m нашей системы. Мы получим равенство

$$(f, \varphi_m) = a_1(\varphi_1, \varphi_m) + a_2(\varphi_2, \varphi_m) + \ldots + a_m(\varphi_m, \varphi_m) + a_{m+1}(\varphi_{m+1}, \varphi_m) + \ldots,$$

из которого в силу того, что $(\varphi_m, \varphi_n) = 0$ при $m \neq n$ и $(\varphi_m, \varphi_m) = 1$, определяется значение коэффициента a_m

$$a_m = (f, \varphi_m) \qquad (m = 1, 2, \ldots).$$

Мы видим, что, как и в обычном трехмерном пространстве (см. начало этого параграфа), коэффициенты a_m равны проекциям вектора f на направления векторов φ_k.

[1] Такое разложение часто встречается в различных вопросах физики при разложении колебаний на гармонические составляющие. См. главу VI (том 2), § 5.

Вспоминая определение скалярного произведения, получаем, что коэффициенты разложения функции $f(x)$ по ортогональной и нормированной системе функций $\varphi_1(x)$, $\varphi_2(x)$, ..., $\varphi_n(x)$, ...

$$f(x) = a_1\varphi_1(x) + a_2\varphi_2(x) + \ldots + a_n\varphi_n(x) + \ldots \tag{17}$$

определяются по формулам

$$a_m = \int_a^b f(t)\,\varphi_m(t)\,dt. \tag{18}$$

В качестве примера рассмотрим ортогональную нормированную тригонометрическую систему функций, приведенную выше:

$$\frac{1}{\sqrt{2\pi}}, \frac{\cos x}{\sqrt{\pi}}, \frac{\sin x}{\sqrt{\pi}}, \frac{\cos 2x}{\sqrt{\pi}}, \frac{\sin 2x}{\sqrt{\pi}}, \ldots$$

Тогда

$$f(x) = \frac{a_0}{2} + \sum_{n=1}^{\infty} a_n \cos nx + b_n \sin nx,$$

где

$$a_0 = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\,dx, \quad a_n = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\cos nx\,dx,$$

$$b_n = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\sin nx\,dx.$$

Мы получили формулу для вычисления коэффициентов разложения функции в тригонометрический ряд в предположении, конечно, что это разложение возможно[1].

Мы установили вид коэффициентов разложения (18) функции $f(x)$ по ортогональной системе функций в предположении, что такое разложение имеет место. Однако бесконечная ортогональная система функций φ_1, φ_2, ..., φ_n, ... может оказаться недостаточной для того, чтобы по ней можно было разложить любую функцию из гильбертова пространства. Чтобы такое разложение было возможно, система ортогональных функций должна удовлетворять дополнительному условию — так называемому условию полноты.

Ортогональная система функций называется *полной*, если к ней нельзя добавить ни одной, не равной тождественно нулю функции, ортогональной ко всем функциям системы.

Легко привести пример неполной ортогональной системы. Для этого возьмем какую-нибудь ортогональную систему, например ту же

[1] О тригонометрических рядах см. также главу XII (том 2), § 7.

систему тригонометрических функций, и исключим одну из функций этой системы, например $\cos x$. Оставшаяся бесконечная система функций

$$1,\ \sin x,\ \cos 2x,\ \sin 2x,\ \ldots,\ \cos nx,\ \sin nx,\ \ldots$$

будет попрежнему ортогональной, но, конечно, не будет полной, так как исключенная нами функция $\cos x$ ортогональна ко всем функциям системы.

Если система функций не полна, то не всякую функцию из гильбертова пространства можно по ней разложить. Действительно, если мы попытаемся разложить по такой системе нулевую функцию $f_0(x)$, ортогональную ко всем функциям системы, то, в силу формул (18), все коэффициенты окажутся равными нулю, в то время как функция $f_0(x)$ не равна нулю.

Имеет место следующая теорема: если задана полная ортогональная и нормированная система функций в гильбертовом пространстве $\varphi_1(x)$, $\varphi_2(x)$, ..., $\varphi_n(x)$, ..., то всякую функцию $f(x)$ можно разложить в ряд по функциям этой системы[1]

$$f(x) = a_1\varphi_1(x) + a_2\varphi_2(x) + \ldots + a_n\varphi_n(x) + \ldots$$

При этом коэффициенты a_n разложения равны проекциям векторов f на элементы ортогональной нормированной системы

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

Имеющаяся в § 2 теорема Пифагора в гильбертовом пространстве позволяет найти интересное соотношение между коэффициентами a_k и функцией $f(x)$. Обозначим через $r_n(x)$ разность между $f(x)$ и суммой первых n членов ее ряда, т. е.

$$r_n(x) = f(x) - [a_1\varphi_1(x) + \ldots + a_n\varphi_n(x)].$$

Функция $r_n(x)$ ортогональна к $\varphi_1(x)$, $\varphi_2(x)$, ..., $\varphi_n(x)$. Действительно, проверим, что она ортогональна к $\varphi_1(x)$, т. е. что $\int_a^b r_n(x)\,\varphi_1(x)\,dx = 0$. Мы имеем

$$\int_a^b r_n(x)\,\varphi_1(x)\,dx = \int_a^b [f(x) - a_1\varphi_1(x) - a_2\varphi_2(x) - \ldots - a_n\varphi_n(x)]\,\varphi_1(x)\,dx =$$

$$= \int_a^b f(x)\,\varphi_1(x)\,dx - a_1\int_a^b \varphi_1^2(x)\,dx.$$[2]

---

[1] Этот ряд относится к своей сумме в том смысле, как это было определено формулой (16).

[2] Остальные интегралы равны нулю, так как функции $\varphi_k(x)$ ортогональны между собой.

Так как $a_1 = \int_a^b f(x)\,\varphi_1(x)\,dx$, а $\int_a^b \varphi_1^2(x)\,dx = 1$, то отсюда следует, что $\int_a^b r_n(x)\,\varphi_1(x)\,dx = 0$.

Итак, в равенстве

$$f(x) = a_1\varphi_1(x) + a_2\varphi_2(x) + \ldots + a_n\varphi_n(x) + r_n(x) \tag{19}$$

отдельные слагаемые правой части ортогональны между собой. Значит, в силу теоремы Пифагора, сформулированной в § 2, квадрат длины вектора $f(x)$ равен сумме квадратов длин слагаемых правой части равенства (19), т. е.

$$\int_a^b f^2(x)\,dx = \int_a^b [a_1\varphi_1(x)]^2\,dx + \ldots + \int_a^b [a_n\varphi_n(x)]^2\,dx + \int_a^b r_n^2(x)\,dx$$

Так как система функций $\varphi_1, \varphi_2, \ldots \varphi_n$ нормирована [равенство (14)], то

$$\int_a^b f^2(x)\,dx = a_1^2 + a_2^2 + \ldots + a_n^2 + \int_a^b r_n^2(x)\,dx. \tag{20}$$

Ряд $\sum_{k=1}^{\infty} a_k\varphi_k(x)$ сходится в среднем. Это значит, что

$$\int_a^b [f(x) - a_1\varphi_1(x) - \ldots - a_n\varphi_n(x)]^2\,dx \to 0.$$

т. е. что

$$\int_a^b r_n^2(x)\,dx \to 0.$$

Но тогда из формулы (20) мы получаем равенство

$$\sum_{k=1}^{\infty} a_k^2 = \int_a^b f^2(x)\,dx\ ^1, \tag{21}$$

утверждающее, что интеграл квадрата функции равен сумме квадратов коэффициентов ее разложения по замкнутой ортогональной системе функций. Условие (21), если оно имеет место для любой функции из гильбертова пространства, называется условием полноты.

Обратим внимание еще на следующее важное обстоятельство. Какие числа a_k могут служить коэффициентами разложения функции из гильбертова пространства? Равенство (21) утверждает, что для этого должен сходиться ряд $\sum_{k=1}^{\infty} a_k^2$. Оказывается, что этого условия достаточно, т. е. для того, чтобы последовательность

[1] Геометрически это значит, что квадрат длины вектора гильбертова пространства равен сумме квадратов его проекций на полную систему взаимно ортогональных направлений.

чисел a_k была последовательностью коэффициентов разложения по ортогональной системе функций [из гильбертова пространства, необходимо и достаточно, чтобы сходился ряд $\sum_{k=1}^{\infty} a_k^2$.

Заметим, что эта существенная теорема имеет место, если определить гильбертово пространство как совокупность всех функций с интегрируемым квадратом в смысле Лебега (см. § 2, стр. 223). Если в гильбертовом пространстве ограничиться только, например, непрерывными функциями, то решение вопроса о том, какие числа a_k могут служить коэффициентами разложения, было бы ненужным образом весьма усложнено.

Приведенные здесь соображения являются лишь одной из причин, приведших к необходимости использовать при определении гильбертова пространства интегралы в обобщенном (по Лебегу) смысле.

§ 4. ИНТЕГРАЛЬНЫЕ УРАВНЕНИЯ

В этом параграфе мы познакомим читателя с одним из важных и исторически одним из первых разделов функционального анализа, а именно с теорией интегральных уравнений, сыгравших также существенную роль в дальнейшем развитии функционального анализа. В развитии теории интегральных уравнений, кроме внутренних потребностей математики [например, краевые задачи для уравнений в частных производных (том 2, глава VI)], имели большое значение различные задачи физики. Наряду с дифференциальными уравнениями в XX в. интегральные уравнения являются одним из важных средств математического изучения различных вопросов физики. В этом параграфе мы изложим некоторые сведения из теории интегральных уравнений. Те факты, которые мы здесь изложим, тесно связаны и в значительной степени возникли (прямо или косвенно) в связи с изучением малых колебаний упругих систем.

Задача о малых колебаниях упругих систем. Вернемся к рассмотренной в § 2 задаче о малых колебаниях. Найдем уравнения, описывающие такие колебания. Для простоты предположим, что мы имеем дело с колебаниями линейной упругой системы. Примерами таких систем могут служить, скажем, струна длины l (рис. 8) или упругий стержень (рис. 9). Предположим, что в положении равновесия наша упругая система расположена по отрезку Ol оси Ox. Приложим в точке x единичную силу. Под действием этой силы все точки системы получат некоторое отклонение. Отклонение, возникшее в точке y (рис. 8), обозначим через $k(x, y)$.

Функция $k(x, y)$ является функцией двух точек: точки x, в которой приложена сила, и точки y, в которой мы измеряем отклонение. Она называется функцией влияния.

Из закона сохранения энергии можно вывести важное свойство функции влияния $k(x, y)$, а именно так называемый закон взаимности: отклонение, возникающее в точке y под действием силы, приложенной

в точке x, равно отклонению, возникающему в точке x под действием той же силы, приложенной в точке y. Другими словами, это значит, что

$$k(x, y) = k(y, x). \tag{22}$$

Найдем, например, функцию влияния для продольных колебаний упругого стержня (на рис. 8 изображались другие, поперечные, смещения). Рассмотрим стержень AB длины l, закрепленный на концах (рис. 9). Приложим в точке C силу f, действующую в направлении B. Под действием этой силы стержень деформируется и точка C сместится в положение C'. Величину смещения точки C обозначим через h. Найдем сначала h. При помощи h мы сможем затем найти смещение в любой точке y. Для этого воспользуемся законом Гука, утверждающим, что сила пропорциональна относительному растяжению (т. е. отношению величины смещения к длине). Аналогичное соотношение имеет место и при сжатии.

Под действием силы f часть стержня AC растягивается. Возникающую при этом силу реакции обозначим через T_1. В то же время часть стержня CB сжимается, порождая силу реакции T_2. В силу закона Гука

$$T_1 = \varkappa \frac{h}{x}, \quad T_2 = \varkappa \frac{h}{l-x},$$

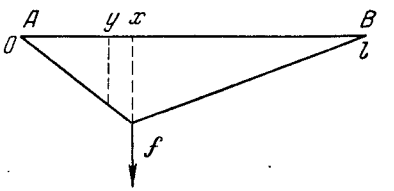


Рис. 8.

где $\varkappa$ — коэффициент пропорциональности, характеризующий упругие свойства стержня. Условие равновесия сил, действующих на точку C, дает нам

$$f = \varkappa \frac{h}{x} + \varkappa \frac{h}{l-x}, \quad \text{т. е.} \quad f = \frac{\varkappa l h}{x(l-x)}.$$

Отсюда

$$h = \frac{f}{\varkappa l} x(l-x).$$

Для того чтобы найти смещение, возникающее в некоторой точке y, лежащей на отрезке AC, т. е. при $y < x$, заметим, что из закона Гука

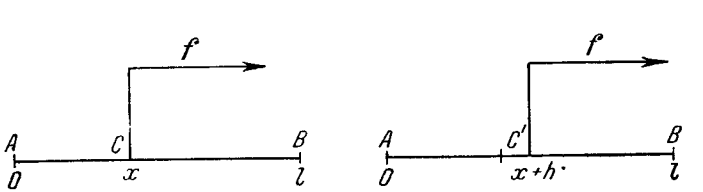


Рис. 9.

следует, что при растяжении стержня относительное растяжение (т. е. отношение смещения точки к расстоянию ее от неподвижного конца)

не зависит от положения точки. Обозначим смещение точки y через k. Тогда, приравняв относительные смещения в точках x и y, получаем

$$\frac{k}{y}=\frac{h}{x},$$

отсюда

$$k=h\frac{y}{x}=\frac{f}{\varkappa l}y(l-x) \quad \text{при} \quad y<x.$$

Аналогично, если точка лежит на отрезке CB $(y>x)$, получаем

$$k=h\frac{l-y}{l-x}=\frac{f}{\varkappa l}x(l-y).$$

Вспоминая, что функция влияния $k(x, y)$ есть отклонение в точке y под действием единичной силы, приложенной в точке x, получаем, что для продольных колебаний упругого стержня функция влияния имеет вид

$$k(x, y)=\begin{cases}\frac{1}{\varkappa l}y(l-x) & \text{при} \quad y<x,\\ \frac{1}{\varkappa l}x(l-y) & \text{при} \quad y>x.\end{cases}$$

Можно было бы более или менее подобным образом найти функцию влияния для струны. Если натяжение струны T, а длина l, то под действием единичной силы, приложенной к точке x, струна приобретает форму, изображенную на рис. 8, и смещение $k(x, y)$ в точке y задается формулой

$$k(x, y)=\begin{cases}\frac{1}{Tl}x(l-y), & \text{при} \quad x<y,\\ \frac{1}{Tl}y(l-x), & \text{при} \quad x>y,\end{cases}$$

совпадающей с выведенной нами функцией влияния для стержня.

Через функцию влияния можно выразить отклонение системы от положения равновесия для того случая, когда на нее действует непрерывно распределенная сила с плотностью $f(y)$. Так как на интервал длины Δy действует сила $f(y)\Delta y$, которую мы приближенно можем считать сосредоточенной в точке y, то под действием этой силы в точке x возникает отклонение $k(x, y)\,f(y)\,\Delta y$. Отклонение под действием всей нагрузки будет приближенно равно сумме

$$\sum k(x, y)\,f(y)\,\Delta y.$$

Переходя к пределу при $\Delta y \to 0$, получаем, что отклонение $u(x)$ в точке x под действием силы $f(y)$, распределенной вдоль системы, задается формулой

$$u(x) = \int_a^b k(x, y) f(y)\, dy. \tag{23}$$

Предположим, что наша упругая система не подвержена действию внешних сил. Если вывести ее из положения равновесия, то она придет в движение. Эти движения называются свободными колебаниями системы.

Напишем теперь при помощи функции влияния $k(x, y)$ уравнение, которому подчиняются свободные колебания рассматриваемой упругой системы. Для этого обозначим через $u(x, t)$ отклонение от положения равновесия в точке x в момент времени t. Тогда ускорение в точке x в момент t равно $\frac{\partial^2 u(x, t)}{\partial t^2}$.

Если ρ — линейная плотность тела, т. е. $\rho\, dy$ — масса элемента длины dy, то по основному закону механики мы получим уравнение движения, заменив в формуле (23) силу $f(y)\, dy$ произведением массы на ускорение $\frac{\partial^2 u(y, t)}{\partial t^2} \rho\, dy$, взятым с обратным знаком.

Таким образом, уравнение свободных колебаний имеет вид

$$u(x, t) = -\int_a^b k(x, y) \frac{\partial^2 u(y, t)}{\partial t^2} \rho\, dy.$$

Важную роль в теории колебаний играют так называемые гармонические колебания упругой системы, т. е. движения, при которых

$$u(x, t) = u(x) \sin \omega t.$$

Они характеризуются тем, что каждая фиксированная точка совершает гармонические колебания (движется по синусоидальному закону) с некоторой частотой ω, причем эта частота одна и та же для всех точек x.

Мы увидим дальше, что каждое свободное колебание может быть составлено из гармонических колебаний.

Подставим

$$u(x, t) = u(x) \sin \omega t$$

в уравнение свободных колебаний и сократим на $\sin \omega t$. Мы получим тогда следующее уравнение для определения функции $u(x)$

$$u(x) = \rho\omega^2 \int_a^b k(x, y) u(y)\, dy. \tag{24}$$

Такое уравнение называется однородным интегральным уравнением относительно функции $u(x)$.

Ясно, что уравнение (24) при любом ω имеет неинтересное для нас решение $u(x) \equiv 0$, отвечающее состоянию покоя. Те значения ω, при которых имеются отличные от нуля решения уравнения (24), называются собственными частотами системы.

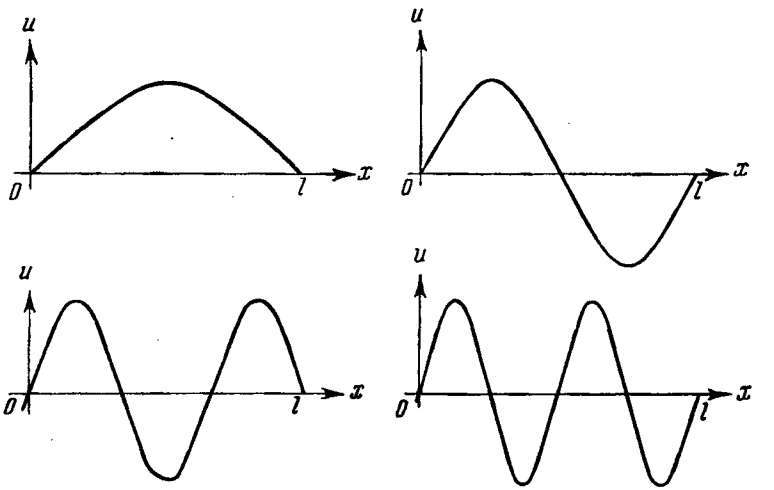


Рис. 10.

Так как не при всяком значении ω имеются отличные от нуля решения, система может совершать свободные колебания лишь с определенными частотами. Наименьшая из них называется основным тоном системы, а остальные — обертонами.

Оказывается, что у каждой системы существует бесконечная последовательность собственных частот, так называемый спектр частот

$$\omega_1, \ \omega_2, \ \ldots, \ \omega_n, \ \ldots$$

Ненулевое решение $u_n(x)$ уравнения (24), отвечающее собственной частоте ω_n, задает нам форму соответствующего собственного колебания.

Так, например, если упругая система представляет собой струну, натянутую между точками O и l и закрепленную в этих точках, то возможные частоты собственных колебаний для такой системы равны

$$a\frac{\pi}{l}, \quad 2a\frac{\pi}{l}, \quad 3a\frac{\pi}{l}, \ \ldots, \ na\frac{\pi}{l}, \ \ldots,$$

где a — коэффициент, зависящий от плотности и натяжения струны, а именно $a = \sqrt{\frac{T}{\rho}}$. Основным тоном является здесь $\omega_1 = a\frac{\pi}{l}$, а обер-

тонами $\omega_2 = 2\omega_1$, $\omega_3 = 3\omega_1$, ..., $\omega_n = n\omega_1$. Формы соответствующих гармонических колебаний задаются равенствами

$$u_n(x) = \sin \frac{n\pi}{l} x$$

и имеют вид, изображенный для $n = 1, 2, 3, 4$ на рис. 10.

Мы рассматривали свободные колебания упругих систем. Если же во время движения на упругую систему действует гармоническая внешняя сила, то, определяя гармонические колебания под действием этой силы, мы придем для функции $u(x)$ к так называемому неоднородному интегральному уравнению

$$u(x) = \rho\omega^2 \int_a^b k(x, y)\, u(y)\, dy + h(x). \tag{25}$$

Свойства интегральных уравнений. Выше мы познакомились с примерами интегральных уравнений

$$f(x) = \lambda \int_a^b k(x, y) f(y)\, dy \tag{26}$$

и

$$f(x) = \lambda \int_a^b k(x, y) f(y)\, dy + h(x), \tag{27}$$

из которых первое получилось при решении задачи о свободных колебаниях упругой системы, а второе — при рассмотрении вынужденных колебаний, т. е. колебаний под воздействием внешней силы.

Неизвестной функцией в этих уравнениях является функция $f(x)$. Заданная функция $k(x, y)$ называется *ядром* интегрального уравнения.

Уравнение (27) называется *неоднородным линейным интегральным уравнением*, а уравнение (26) — *однородным*. Оно получается из неоднородного при $h(x) = 0$.

Ясно, что однородное уравнение всегда имеет нулевое решение, т. е. решение $f(x) = 0$. Между решениями неоднородного и однородного интегральных уравнений существует тесная связь. Укажем для примера на следующую теорему: если однородное интегральное уравнение имеет лишь нулевое решение, то соответствующее неоднородное уравнение разрешимо для всякой функции $h(x)$.

Если для некоторого значения λ однородное уравнение имеет решение $f(x)$, не равное тождественно нулю, то это значение λ называется *собственным значением*, а соответствующее решение $f(x)$ — *собственной функцией*. Мы видели выше, что когда интегральное уравнение описывает свободное колебание упругой системы, то собственные значения тесно связаны с частотами колебаний системы (а именно $\lambda = \rho\omega^2$). Собственные же функции задают форму соответствующих гармонических колебаний.

Для задачи о колебаниях из закона сохранения энергии следовало, что

$$k(x, y) = k(y, x). \tag{28}$$

Ядро, удовлетворяющее условию (28), называется *симметрическим*.

Для уравнения с симметрическим ядром собственные функции и собственные значения обладают рядом важных свойств. Оказывается, что у такого уравнения всегда существует последовательность действительных собственных значений

$$\lambda_1, \lambda_2, \ldots, \lambda_n, \ldots$$

Каждому собственному значению отвечает одна или несколько собственных функций. При этом собственные функции, отвечающие различным собственным значениям, всегда ортогональны между собою[1].

Таким образом, для всякого интегрального уравнения с симметрическим ядром система собственных функций есть ортогональная система функций. Возникает вопрос о том, когда эта система полна, т. е. когда можно всякую функцию из гильбертова пространства разложить в ряд по системе собственных функций интегрального уравнения. В частности, если уравнение

$$\int_a^b k(x, y) f(y)\, dy = 0 \tag{29}$$

удовлетворяется лишь при $f(y) \equiv 0$, то система собственных функций интегрального уравнения

$$\lambda \int_a^b k(x, y) f(y)\, dy = f(x)$$

является полной ортогональной системой[2].

Таким образом, всякую функцию $f(x)$ с интегрируемым квадратом можно в этом случае разложить в ряд по собственным функциям. Рассматривая те или иные интегральные уравнения, мы получаем общий

---

[1] Это последнее утверждение будет доказано в следующем параграфе (стр. 243).

[2] В том случае, когда $k(x, y)$ есть функция влияния для упругой системы, условие (29) приобретает простой физический смысл. В самом деле [см. формулу (23)], мы видели, что под действием распределенной вдоль системы силы $f(y)$ отклонение системы от положения равновесия выражается формулой $u(x) = \int_a^b k(x, y) f(y)\, dy$. Таким образом, условие (29) означает, что всякая отличная от нуля сила выводит систему из положения равновесия.

и сильный метод для доказательства замкнутости различных важных ортогональных систем, т. е. разложимости функций в ряд по ортогональным функциям. Этим методом можно доказать полноту системы тригонометрических функций, цилиндрических функций, сферических функций и многих других важных систем функций.

То обстоятельство, что произвольную функцию можно разложить в ряд по собственным функциям, в случае колебаний означает, что любое колебание может быть разложено на сумму гармонических. Такое разложение представляет собой метод, широко применяемый при решении задач о колебаниях в различных областях механики и физики (колебания упругих тел, акустические колебания, электромагнитные волны и т. д.).

Развитие теории линейных интегральных уравнений явилось толчком для создания общей теории линейных операторов, в которую теория линейных интегральных уравнений органически входит. В последние несколько десятилетий общие методы теории линейных операторов сильно способствовали дальнейшему развитию теории интегральных уравнений.

§ 5. ЛИНЕЙНЫЕ ОПЕРАТОРЫ И ДАЛЬНЕЙШЕЕ РАЗВИТИЕ ФУНКЦИОНАЛЬНОГО АНАЛИЗА

В предыдущем параграфе мы видели, что задачи о колебаниях упругой системы приводятся к вопросу о разыскании собственных значений и собственных функций интегральных уравнений. Заметим, что эти задачи можно также свести к отысканию собственных значений и собственных функций линейных дифференциальных уравнений[1]. К задачам вычисления собственных значений и собственных функций линейных дифференциальных или интегральных уравнений сводятся и многие другие физические задачи.

Укажем еще один пример. В современной радиотехнике широко используются для передачи электромагнитных колебаний высоких частот так называемые волноводы, т. е. полые металлические трубы, внутри которых распространяются электромагнитные волны. Известно, что по волноводам могут распространяться электромагнитные колебания не слишком большой длины волны. Отыскание критической длины волны сводится к задаче на собственные значения некоторого дифференциального уравнения.

Кроме того, задачи на собственные значения встречаются в линейной алгебре, в теории обыкновенных дифференциальных уравнений, в задачах на устойчивость и т. д.

Возникла необходимость в рассмотрении всех этих сходных между собой вопросов с единой общей точки зрения. Такой точкой зрения

[1] См. том 2, главу VI, § 5.

явилась общая теория линейных операторов. Многие вопросы, относящиеся к собственным функциям и собственным значениям в различных конкретных задачах, удалось понять до конца только в свете общей теории операторов. Таким образом, в этом и ряде других направлений общая теория операторов сама оказалась весьма плодотворным орудием исследования в породивших ее областях.

В дальнейшем развитии теории операторов существеннейшим этапом явилась квантовая механика, широко использующая методы теории операторов. Основным математическим аппаратом квантовой механики и является теория так называемых самосопряженных операторов. Постановка математических задач, возникающих в квантовой механике, явилась и является важнейшим стимулом для дальнейшего развития функционального анализа.

Операторная точка зрения на дифференциальные и интегральные уравнения оказалась чрезвычайно полезной также для развития практических приближенных методов решения этих уравнений.

Основные понятия теории операторов. Перейдем к изложению основных определений и фактов, относящихся к теории операторов.

В анализе мы уже встречались с понятием функции. В простейшем виде это было соответствие, которое каждому числу x (значению независимой переменной) ставило в соответствие число y (значение функции). При дальнейшем развитии анализа возникла потребность в рассмотрении соответствий более общего типа.

Такие более общие соответствия рассматриваются, например, в вариационном исчислении (том 2, глава VIII), где каждой функции сопоставляется число. Если каждой функции ставится в соответствие число, то мы говорим, что нам задан функционал. Примером функционала может служить сопоставление каждой функции $y=f(x)$ $(a\leqslant x\leqslant b)$ длины дуги изображаемой ею кривой. Мы получим другой пример функционала, если каждой функции $y=f(x)$ $(a\leqslant x\leqslant b)$ поставим в соответствие ее определенный интеграл $\int\limits_a^b f(x)\,dx$.

Если считать $f(x)$ точкой бесконечномерного пространства, то функционал есть не что иное, как функция от точки бесконечномерного пространства. С этой точки зрения в вариационном исчислении изучаются задачи об отыскании максимумов и минимумов функции от точки бесконечномерного пространства.

Для того чтобы определить, что означает непрерывный функционал, нужно определить, что означает близость двух точек бесконечномерного пространства. В § 2 мы задавали расстояние между двумя функциями [точками бесконечномерного пространства $f(x)$ и $g(x)$] как $\sqrt{\int\limits_a^b [f(x)-g(x)]^2\,dx}$. Такой способ задания расстояния в бесконечно-

мерном пространстве часто употребляется, но не является, конечно, единственно возможным. В других вопросах могут оказаться лучшими другие способы задания расстояния между функциями. Укажем, например, на задачи теории приближения функций (см. том 2, главу XII, § 3), где расстояние между функциями, характеризующее меру близости двух функций $f(x)$ и $g(x)$, задается, например, формулой

$$\max |f(x) - g(x)|.$$

Другие способы задания расстояния между функциями употребляются при изучении функционалов в вариационном исчислении.

Различные способы задания расстояния между функциями приводят нас к различным бесконечномерным пространствам.

Таким образом, разные бесконечномерные (функциональные) пространства различаются между собой запасом функций и определением расстояния между ними. Так, например, если взять совокупность всех функций с интегрируемым квадратом и определить расстояние как $\sqrt{\int_a^b [f(x) - g(x)]^2 dx}$, то мы придем к введенному в § 2 гильбертову пространству; если же взять совокупность всех непрерывных функций и определить расстояние как $\max |f(x) - g(x)|$, то мы получим так называемое пространство (C).

При рассмотрении интегральных уравнений мы сталкиваемся с выражениями вида

$$g(x) = \int_a^b k(x, y) f(y)\, dy.$$

При заданном ядре $k(x, y)$ написанное равенство указывает закон, по которому каждой функции $f(x)$ ставится в соответствие другая функция $g(x)$.

Такого рода соответствие, относящее одной функции f другую функцию g, называется *оператором*.

Будем говорить, что нам задан линейный оператор A в гильбертовом пространстве, если дан закон, по которому каждой функции f ставится в соответствие функция g. Соответствие может быть задано и не для всех функций гильбертова пространства. В этом случае множество тех функций f, для которых существует функция $g = Af$, называется *областью определения* оператора A (аналогично области задания функции в обычном анализе). Само соответствие обозначается обычно так:

$$g = Af. \tag{30}$$

Линейность оператора означает, что сумме функций f_1 и f_2 ставится в соответствие сумма Af_1 и Af_2, а произведению функции f на число λ ставится в соответствие функция λAf, т. е.

$$A(f_1+f_2)=Af_1+Af_2 \tag{31}$$

и

$$A(\lambda f)=\lambda Af. \tag{32}$$

Иногда от линейных операторов требуют также непрерывности, т. е. требуют, чтобы при сходимости последовательности функций f_n к функции f последовательность Af_n сходилась к функции Af.

Приведем примеры линейных операторов.

1°. Поставим в соответствие каждой функции $f(x)$ функцию $g(x)=\int_a^x f(t)\,dt$, т. е. неопределенный интеграл от функции f. Линейность этого оператора вытекает из обычных свойств интеграла, т. е. из того, что интеграл от суммы равен сумме интегралов и постоянный множитель можно выносить за знак интеграла.

2°. Поставим в соответствие каждой дифференцируемой функции $f(x)$ ее производную $f'(x)$. Этот оператор обозначается обычно буквой D, т. е.

$$f'(x)=Df(x).$$

Заметим, что этот оператор определен не для всех функций из гильбертова пространства, а только для функций, имеющих производную, и притом принадлежащую гильбертову пространству. Эти функции составляют, как уже сказано, область определения данного оператора.

3°. Примеры 1° и 2° были примерами линейных операторов в бесконечномерном пространстве. С примерами линейных операторов в конечномерных пространствах мы встречались в других главах этой книги. Так, в главе III (том 1) изучались аффинные преобразования. Если аффинное преобразование плоскости или пространства оставляет начало координат на месте, то оно является примером линейного оператора в двумерном, соответственно трехмерном, пространстве. В главе XVI вводились линейные преобразования n-мерного пространства, являющиеся линейными операторами в n-мерном пространстве.

4°. В интегральных уравнениях мы уже по существу встречались с весьма важным и имеющим широкое применение в анализе классом линейных операторов в функциональном пространстве — так называемыми интегральными операторами. Зададим некоторую определенную функцию $k(x, y)$. Тогда формула

$$g(x)=\int_a^b k(x,y)f(y)\,dy$$

ставит в соответствие каждой функции f некоторую функцию g. Символически мы можем это преобразование записать следующим образом:

$$g = Af.$$

Оператор A называется в этом случае интегральным оператором. Можно было бы привести еще много важных примеров интегральных операторов.

В § 4 мы говорили о неоднородном интегральном уравнении

$$f(x) = \lambda \int_a^b k(x, y) f(y)\, dy + h(x).$$

В обозначениях теории операторов это уравнение перепишется так:

$$f = \lambda Af + h, \tag{33}$$

где λ — заданное число, h — заданная функция (вектор бесконечного пространства), f — искомая функция. Однородное уравнение в тех же обозначениях переписывается следующим образом:

$$f = \lambda Af. \tag{34}$$

Классические теоремы об интегральных уравнениях, как, например, сформулированная в § 4 теорема о связи между разрешимостью неоднородного и соответствующего однородного интегральных уравнений, справедливы не для всякого операторного уравнения. Однако можно указать некоторые общие условия, накладываемые на оператор A, при которых эти теоремы верны.

Эти условия формулируются в топологических терминах и состоят в том, чтобы оператор A переводил единичную сферу (т. е. совокупность векторов, длины которых не превосходят единицы) в компактное множество.

Собственные значения и собственные векторы операторов. Задача о собственных значениях и собственных функциях интегрального уравнения, к которой нас привели задачи о колебаниях, формулировалась следующим образом: найти значения λ, при которых имеются отличные от нуля функции f, удовлетворяющие уравнению

$$f(x) = \lambda \int_a^b k(x, y) f(y)\, dy.$$

Как и ранее, это уравнение может быть переписано так:

$$f = \lambda Af$$

или

$$Af = \frac{1}{\lambda} f. \tag{35}$$

Будем теперь понимать под A произвольный линейный оператор. Тогда вектор f, удовлетворяющий равенству (35), называется собственным вектором оператора A, а число $\frac{1}{\lambda}$ — соответствующим собственным значением.

Так как вектор $\frac{1}{\lambda} f$ по направлению совпадает с вектором f (отличается от f лишь численным множителем), то задача разыскания собственных векторов может быть еще сформулирована как задача о разыскании ненулевых векторов f, не меняющих своего направления при преобразовании A.

Такая точка зрения на проблему о собственных значениях позволяет объединить задачу о собственных значениях интегральных уравнений (если A — интегральный оператор), дифференциальных уравнений (если A — дифференциальный оператор) и задачу о собственных значениях в линейной алгебре [если A является линейным преобразованием в конечномерном пространстве; см. главу VI (том 2) и главу XVI]. Для случая трехмерного пространства эта проблема встречается при отыскании так называемых главных осей эллипсоида.

В случае интегральных уравнений ряд важных свойств собственных функций и собственных значений (например, вещественность собственных значений, ортогональность собственных функций и т. д.) является следствием симметричности ядра, т. е. равенства $k(x, y) = k(y, x)$.

Для произвольного линейного оператора A в гильбертовом пространстве аналогом этого свойства является так называемая самосопряженность оператора.

Условие самосопряженности оператора A в общем случае заключается в том, что для любых двух элементов f_1 и f_2 имеет место равенство

$$(Af_1, f_2) = (f_1, Af_2),$$

где (Af_1, f_2) означает скалярное произведение вектора Af_1 на вектор f_2.

Требование самосопряженности оператора в задачах механики является обычно следствием закона сохранения энергии. Поэтому оно удовлетворяется для операторов, связанных, например, с колебаниями, при которых не имеет места потеря (диссипация) энергии.

Большинство операторов, встречающихся в квантовой механике, также являются самосопряженными.

Проверим, что интегральный оператор с симметрическим ядром $k(x, y)$ является самосопряженным. Действительно, в этом случае Af_1 есть функция $\int_a^b k(x, y) f_1(y)\, dy$. Поэтому скалярное произведение (Af_1, f_2), равное интегралу от произведения этой функции на f_2, задается формулой

$$(Af_1, f_2) = \int_a^b \int_a^b k(x, y) f_1(y) f_2(x)\, dy\, dx.$$

Аналогично

$$(f_1, Af_2) = \int_a^b \int_a^b k(x, y) f_2(y) f_1(x)\, dy\, dx.$$

Равенство $(Af_1, f_2) = (f_1, Af_2)$ есть непосредственное следствие симметричности ядра $k(x, y)$.

Произвольные самосопряженные операторы обладают рядом важнейших свойств, полезных при применении этих операторов к решению разного рода задач. Именно, оказывается, что собственные значения самосопряженного линейного оператора всегда действительны и собственные функции, отвечающие различным собственным значениям, ортогональны между собой.

Докажем, например, последнее утверждение. Пусть λ_1 и λ_2 — два различных собственных значения оператора A, а f_1 и f_2 — соответствующие им собственные векторы. Это значит, что

$$\begin{aligned} Af_1 &= \lambda_1 f_1, \\ Af_2 &= \lambda_2 f_2. \end{aligned} \tag{36}$$

Умножим скалярно первое из равенств (36) на f_2, а второе — на f_1. Мы имеем

$$\begin{aligned} (Af_1, f_2) &= \lambda_1 (f_1, f_2), \\ (Af_2, f_1) &= \lambda_2 (f_2, f_1). \end{aligned} \tag{37}$$

Так как оператор A самосопряженный, то $(Af_1, f_2) = (Af_2, f_1)$. Вычитая из первого равенства (37) второе, получаем

$$0 = (\lambda_1 - \lambda_2)(f_1, f_2).$$

Так как $\lambda_1 \neq \lambda_2$, то $(f_1, f_2) = 0$, т. е. собственные векторы f_1 и f_2 ортогональны.

Исследование самосопряженных операторов внесло ясность во многие конкретные вопросы и задачи, связанные с теорией собственных значений.

Остановимся подробнее на одной из них, а именно на задаче о разложении по собственным функциям в случае непрерывного спектра.

Чтобы пояснить, что значит непрерывный спектр, обратимся снова к классическому примеру колебаний струны. Мы указывали выше, что для струны длины l собственные частоты колебаний могут принимать последовательность значений

$$a\frac{\pi}{l},\ 2a\frac{\pi}{l},\ \ldots,\ na\frac{\pi}{l},\ \ldots$$

Отложим на числовой оси $O\lambda$ точки этой последовательности. Если увеличивать длину струны l, то расстояние между любыми двумя соседними точками последовательности будет уменьшаться, и они будут всё более плотно заполнять числовую ось. В предельном случае, когда $l \to \infty$, т. е. для бесконечной струны, собственные частоты заполнят числовую полуось $\lambda \geqslant 0$. В этом случае говорят, что система имеет непрерывный спектр.

Мы уже говорили, что разложение в ряд по собственным функциям для струны длины l есть разложение в ряд по синусам и косинусам $n\frac{\pi}{l}x$, т. е. в тригонометрический ряд

$$f(x) = \frac{a_0}{2} + \sum a_n \cos n\frac{\pi}{l}x + b_n \sin n\frac{\pi}{l}x.$$

Для случая бесконечной струны снова можно показать, что более или менее произвольную функцию можно разложить по синусам и косинусам. Однако, поскольку собственные частоты распределены теперь по числовой прямой непрерывно, это разложение будет не разложением в ряд, а разложением в так называемый интеграл Фурье

$$f(x) = \int_{-\infty}^{+\infty} [A(\lambda)\cos\lambda x + B(\lambda)\sin\lambda x]\, d\lambda.$$

Разложение в интеграл Фурье было давно известно и широко использовалось уже в XIX в. при решении различных задач математической физики.

Однако в более общих задачах с непрерывным спектром[1] многие вопросы, относящиеся к разложению функций по собственным функциям, не были выяснены. Только создание общей теории самосопряженных операторов внесло необходимую ясность в эти вопросы.

Отметим еще один круг классических задач, получивших свое разрешение на основе общей теории операторов. К этим задачам относится рассмотрение колебаний при наличии диссипации (рассеяния) энергии.

В этом случае мы снова можем искать свободные колебания системы в виде $u(x)\varphi(t)$. Однако, в отличие от случая колебаний без диссипации энергии, функция $\varphi(t)$ не есть просто $\cos\omega t$, а имеет вид $e^{-kt}\cos\omega t$, где $k > 0$. Таким образом, соответствующее решение имеет вид $u(x)e^{-kt}\cos\omega t$. Каждая точка x в этом случае снова совершает колебания (с частотой ω), однако колебания являются затухающими, так как

[1] Примерами могут служить колебания неоднородной упругой среды, а также многие задачи квантовой механики.

при $t \to \infty$ амплитуда этих колебаний, содержащая множителем e^{-kt}, стремится к нулю.

Удобно записывать собственные колебания системы в комплексной форме: $u(x)e^{i\lambda t}$, где при отсутствии трения число λ действительно, а при наличии трения комплексно.

Задача о колебаниях системы с диссипацией энергии опять приводит к задаче о собственных значениях, но уже не для самосопряженных операторов. Для них характерно наличие комплексных собственных значений, свидетельствующих о затухании свободных колебаний.

Используя методы теории операторов в соединении с методами теории аналитических функций, М. В. Келдыш в 1950—1951 гг. изучил этот класс задач, доказав для него полноту системы собственных функций.

Связь функционального анализа с другими разделами математики и квантовой механикой. Мы уже упоминали выше, что создание квантовой механики явилось решающим моментом в развитии функционального анализа.

Так же, как возникновение дифференциального и интегрального исчисления в XVIII в. диктовалось потребностями механики и классической физики, развитие функционального анализа происходило и происходит под сильнейшим влиянием современной физики, главным образом квантовой механики. Основным математическим аппаратом квантовой механики являются разделы математики, относящиеся по существу к функциональному анализу. Мы лишь вкратце укажем на имеющиеся здесь связи, поскольку изложение основ квантовой механики выходит за рамки этой книги.

В квантовой механике состояние системы задается при ее математическом описании вектором гильбертова пространства. Такие величины, как энергия, импульс, момент количества движения, исследуются с помощью самосопряженных операторов. Например, возможные энергетические уровни электрона в атоме вычисляются как собственные значения оператора энергии. Разности этих собственных значений дают частоты излучаемого атомом света, определяя, таким образом, структуру спектра излучения данного вещества. Соответствующие состояния электрона описываются при этом как собственные функции оператора энергии.

Решение задач квантовой механики часто требует вычисления собственных значений различных (обычно дифференциальных) операторов. В сколько-нибудь сложных случаях точное решение этих задач оказывается практически невозможным. Для приближенного решения этих задач широко используется так называемая теория возмущений, позволяющая по уже известным собственным значениям и функциям некоторого самосопряженного оператора A находить собственные значения

мало отличающегося от него оператора A_1. Отметим, что теория возмущений далека от полного математического обоснования, которое является интересной и важной математической проблемой.

Независимо от приближенного определения собственных значений часто можно много сказать о данной задаче при помощи качественного исследования. Это исследование в задачах квантовой механики проводится на основе имеющихся в данной задаче симметрий. Примерами таких симметрий могут служить свойства симметрии кристаллов, сферическая симметрия в атоме, симметрия относительно отражений и др. Так как симметрии образуют группу (см. главу XX), то групповые методы (так называемая теория представлений групп) дают возможность ответить без вычислений на ряд вопросов. Укажем, например, на классификацию атомных спектров, ядерные превращения и другие вопросы.

Таким образом, квантовая механика широко использует математический аппарат теории самосопряженных операторов. В то же время продолжающееся в настоящее время развитие квантовой механики приводит к дальнейшему развитию теории операторов, ставя перед этой теорией новые задачи.

Влияние квантовой механики, а также и внутриматематическое развитие функционального анализа, привело к тому, что в последние годы алгебраические проблемы и методы стали играть значительную роль в функциональном анализе. Это усиление алгебраических тенденций в современном анализе не лишне сопоставить с возросшим значением алгебраических методов в современной теоретической физике по сравнению с методами физики XIX в.

В заключение подчеркнем еще раз, что функциональный анализ является одним из интенсивно развивающихся разделов современной математики. Его связи и применения в современной физике, дифференциальных уравнениях, приближенных вычислениях и использование общих методов, выработанных в алгебре, топологии, теории функций действительного переменного и т. п. делают функциональный анализ одним из основных узлов современной математики.

ЛИТЕРАТУРА

Университетские учебники

Колмогоров А. Н. и Фомин С. В. Элементы теории функций и функционального анализа, вып. 1. Метрические и нормированные пространства. Изд-во МГУ, 1954.

Курант Р. и Гильберт Д. Методы математической физики, т. I. Гостехиздат, 1951.

В книге, наряду с другим материалом, изложены связи теории колебаний с теорией собственных значений.

Люстерник Л. А. и Соболев В. И. Элементы функционального анализа. Гостехиздат, 1951.

Рисс Ф. и Секефальви-Надь Б. Лекции по функциональному анализу. ИЛ, 1954.

Книга, требующая от читателя значительной математической подготовки.

Шилов Г. Е. Введение в теорию линейных пространств. Изд. 2, Гостехиздат, 1956.

Для чтения этой книги достаточно знания основ анализа и аналитической геометрии.

Литература, относящаяся к вопросам, затронутым в конце главы

Ван дер Варден Б. Метод теории групп в квантовой механике. Харьков, 1938.

Гельфанд И. М. и Шапиро З. Я. Представления группы вращений трехмерного пространства и их применения. Успехи матем. наук, 7, № 1, 3—117, 1952.

Ландау Л. О. и Лифшиц Е. А. Квантовая механика. Гостехиздат, 1948.

Глава XX

ГРУППЫ И ДРУГИЕ АЛГЕБРАИЧЕСКИЕ СИСТЕМЫ

§ 1. ВВЕДЕНИЕ

В главе IV (том 1), посвященной алгебре многочленов, уже шла речь об основных путях развития алгебры, ее месте среди других математических дисциплин, об изменениях во взглядах на самый предмет алгебры. Цель настоящей главы заключается в том, чтобы дать читателю представление о тех новых алгебраических теориях, которые, возникнув еще в прошлом веке, достигли полного развития в текущем столетии и оказали большое влияние на современные математические исследования.

Современная алгебра, так же как классическая, есть учение о действиях, о правилах вычислений. Но она не ограничивается изучением свойств действий над числами, а стремится изучать свойства действий над элементами все более общей природы. Эта тенденция диктуется потребностями практики. Так, в механике складываются силы, скорости, повороты. В линейной алгебре (см. главу XVI), идеи и методы которой находят широкое применение в практических расчетах, областью действий являются матрицы, линейные преобразования, векторы n-мерного пространства.

Особо выдающуюся роль в современной алгебре играет теория групп, которой и будет посвящена бóльшая часть этой главы. Из других алгебраических теорий мы остановимся на теории гиперкомплексных систем, являющейся необходимым и важным этапом в историческом процессе развития понятия числа. Этими двумя теориями, конечно, далеко не исчерпывается содержание современной алгебры, но ее идеи и методы освещаются ими достаточно ясно.

Теория групп возникла из необходимости найти аппарат для изучения таких важных закономерностей реального мира, как закономерность симметрии.

Познание свойств симметрии каких-либо геометрических тел или других математических и физических объектов иногда дает ключ к выяснению строения этих тел и объектов. Однако, несмотря на всю наглядность понятия симметрии, точная и общая формулировка того, что такое симметрия, и в особенности количественный учет свойств симметрии требуют использования аппарата теории групп.

Теория групп возникла сравнительно давно: в конце XVIII и начале XIX в. Первоначально она развивалась лишь как вспомогательный аппарат для задачи о решении уравнений высших степеней в радикалах. Это было вызвано тем, что именно в указанной задаче впервые было замечено, что свойства равноправности, симметрии корней уравнения являются основными для решения всей задачи. В течение XIX и XX вв. важная роль закономерностей симметрии выявилась во многих других разделах науки: геометрии, кристаллографии, физике, химии. Благодаря этому методы и результаты теории групп получили широкое распространение. Поскольку каждая область приложений ставила перед теорией групп свои особенные задачи, рост числа этих областей оказывал и обратное воздействие, вызывая развитие новых отделов теории групп, приведшее к тому, что современная теория групп, являясь единой по своим основным понятиям, фактически распадается на ряд более или менее самостоятельных дисциплин: общая теория групп, теория конечных групп, теория непрерывных групп, дискретные группы преобразований, теория представлений и характеров групп. Постепенно развиваясь, методы и понятия теории групп оказались важными не только для изучения закономерностей симметрии, но и для решения многих других вопросов.

В настоящее время понятие группы стало одним из важнейших обобщающих понятий современной математики, а теория групп заняла видное место среди математических дисциплин. Выдающийся вклад в развитие теории групп и ее приложений внесли Е. С. Федоров, О. Ю. Шмидт, Л. С. Понтрягин. Исследования советских математиков в области теории групп занимают ведущее место и в современном развитии этой теории.

§ 2. СИММЕТРИЯ И ПРЕОБРАЗОВАНИЯ

Простейшие виды симметрии. Начнем с напоминания простейших видов симметрии, знакомых читателю из повседневной жизни. Одним из таких видов симметрии является зеркальная симметрия геометрических тел или симметрия относительно плоскости.

Точка A пространства называется симметричной точке B относительно плоскости α (рис. 1), если плоскость пересекает отрезок AB в его середине перпендикулярно этому отрезку. Говорят также, что точка B является зеркальным образом точки A относительно плоскости α. Геометрическое тело называется симметричным относительно плоскости, если эта плоскость разбивает тело на две части, из которых каждая является зеркальным отражением другой относительно данной плоскости. Сама плоскость называется в этом случае плоскостью симметрии тела. Зеркальная симметрия очень распространена в природе. Например, форма человеческого тела, форма тела зверей, птиц имеет обычно плоскость симметрии.

Симметрия относительно прямой определяется аналогичным образом. Говорят, что точки A, B лежат симметрично относительно прямой, если эта прямая пересекает отрезок AB в его средней точке и перпендикулярна AB (рис. 2). Геометрическое тело называется симметричным относительно прямой или имеющим эту прямую своей осью симметрии 2-го порядка, если для каждой точки тела симметричная точка также принадлежит телу.

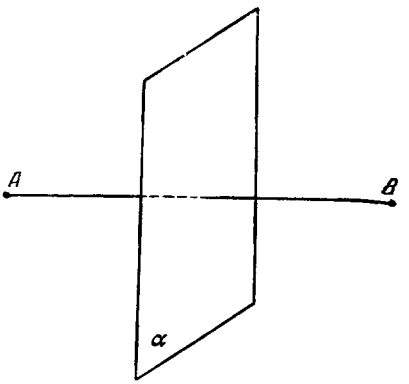


Рис. 1.

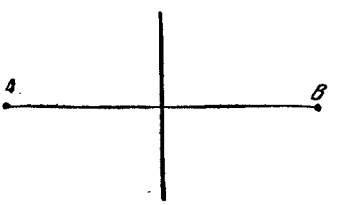


Рис. 2.

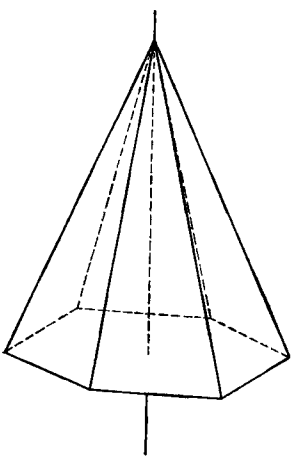


Рис. 3.

Тело, имеющее ось симметрии 2-го порядка, совмещается с собой при повороте тела вокруг этой оси на половину полного поворота, т. е. на угол в 180°.

Понятие оси симметрии естественным образом обобщается. Прямая называется осью симметрии порядка n для данного тела, если это тело совмещается с собой при повороте вокруг оси на угол $\frac{1}{n}$ 360°. Так, правильная пирамида, основанием которой является правильный n-угольник, имеет осью симметрии порядка n прямую, соединяющую вершину пирамиды с центром основания (рис. 3).

Прямая называется осью вращения тела, если тело совмещается с собой при повороте вокруг оси на любой угол. Так, оси цилиндра и конуса, любой диаметр шара суть их оси вращения. Ось вращения является вместе с тем осью симметрии любого порядка.

Наконец, важным типом симметрии является также симметрия относительно точки или центральная симметрия. Точки A и B называются симметричными относительно центра O, если отрезок, соединяющий точки A и B, делится точкой O пополам. Тело называется симметричным относительно центра O, если все его точки распадаются на пары точек, симметричных относительно O. Примерами центрально-

симметричных тел могут служить шар и куб, центры которых являются их центрами симметрии (рис. 4).

Знание всех плоскостей, осей и центров симметрии тела дает довольно полное представление о свойствах его симметрии.

Однако понятие симметрии имеет смысл не только в применении к геометрическим фигурам. Например, имеет совершенно ясный смысл утверждение, что в многочлене $x_1^3+x_2^3+x_3^3+x_4^3$ переменные x_1, x_2, x_3, x_4 входят симметрично, а в многочлене $x_1^3+x_2^2+x_3^2+x_4^3$ входят симметрично переменные x_1 и x_4, x_2 и x_3, в то время как, например, переменные x_1 и x_2 играют различную роль. Число таких примеров можно легко увеличить. Это заставляет поставить важный вопрос: что же такое симметрия в общем случае и как можно математически учитывать отношения симметрии? Оказывается, точный ответ на этот вопрос связан с понятием преобразования, которое уже много раз встречалось в этой книге, начиная с самых первых ее глав. Чтобы иметь возможность дать общее определение симметрии, охватывающее такие разнородные случаи, как симметрия пространственных тел и симметрия многочленов, необходимо и понятие преобразования сформулировать в очень общем виде.

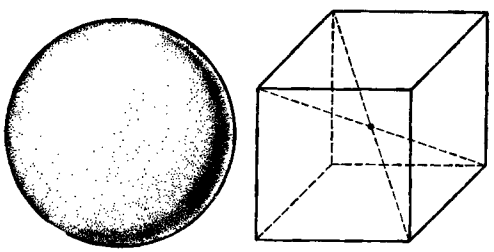


Рис. 4.

Преобразования. Пусть M обозначает конечную или бесконечную совокупность совершенно произвольных объектов. Например, M может быть совокупностью чисел 1, 2, . . . , n, совокупностью независимых переменных x_1, x_2, x_3, x_4, множеством всех точек плоскости. Если каждому элементу множества M поставлен в соответствие вполне определенный элемент того же множества, то говорят, что задано преобразование множества M. Каждое преобразование конечного множества M можно задать посредством таблицы, состоящей из двух строк: в верхней строке пишутся в произвольном порядке обозначения элементов множества M и под каждым из них записывается обозначение соответствующего ему элемента. Например, таблица $\begin{pmatrix}1&2&3&4\\2&3&2&1\end{pmatrix}$ обозначает преобразование совокупности чисел 1, 2, 3, 4, при котором числа 1, 2, 3, 4 переходят соответственно в числа 2, 3, 2, 1. Располагая в верхней строке числа 1, 2, 3, 4 в порядке 3, 4, 1, 2, мы можем то же самое преобразование записать и таблицей $\begin{pmatrix}3&4&1&2\\2&1&2&3\end{pmatrix}$.

Если множество M бесконечно, но имеется возможность перечислить (перенумеровать) его элементы, расположив их обозначения в одну

строку (например, если M есть множество всех чисел натурального ряда 1, 2, 3, ...), то преобразование можно задавать аналогичным образом.

Изучая преобразования, нужно ввести для них целесообразные обозначения. Будем обозначать преобразования просто буквами A, B и т. д., причем если какое-либо преобразование множества M будет обозначено буквой A, то через mA, где m — произвольный элемент из M, будет обозначаться образ элемента m, т. е. тот элемент, в который переходит m при преобразовании A. Пусть, например, $A=\begin{pmatrix}1&2&3&4\\2&3&2&1\end{pmatrix}$, тогда

$$1A=2,\quad 2A=3,\quad 3A=2,\quad 4A=1.$$

Укажем несколько преобразований, играющих важную роль в геометрии.

Выберем в пространстве какую-либо прямую a и поставим в соответствие каждой точке P пространства точку Q, получаемую путем поворота точки P вокруг оси a на фиксированный угол φ (рис. 5). Тем самым мы определили преобразование совокупности всех точек пространства, называемое поворотом пространства на угол φ вокруг оси a.

Заметим, что слово «поворот» в механике означает некоторый процесс, в результате которого точки тела переходят в новое положение.

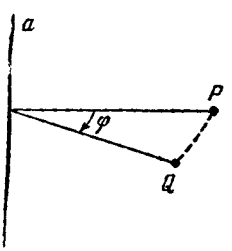


Рис. 5.

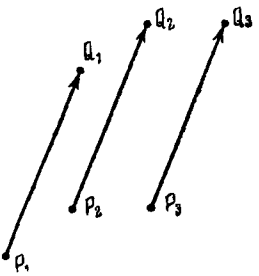


Рис. 6.

Здесь же термин «поворот» употребляется в смысле преобразования пространства. При этом отвлекаются от самого процесса движения и рассматривают только его конечный результат — соответствие начального и конечного положения точек.

Другим важным преобразованием пространства является параллельный перенос всех точек в данном направлении на заданное расстояние. Из рис. 6, на котором для произвольных точек P_1, P_2, P_3 указаны соответствующие точки Q_1, Q_2, Q_3, видно, что, зная при параллельном переносе соответственную точку лишь для одной точки пространства, можно найти соответственные точки для всех других точек пространства.

Выше были определены понятия плоскости, оси и центра симметрии пространственной фигуры. Каждому из этих понятий отвечает определенное преобразование пространства: отражение относительно плоскости, вращение относительно прямой и отражение относительно центра. Например, отражение относительно плоскости есть преобразование, при котором каждой точке пространства ставится в соответствие точка, симметричная относительно этой плоскости. Аналогично определяется вращение относительно прямой и отражение относительно центра.

Мы говорили о преобразованиях пространства. Соответствующие преобразования плоскости: поворот плоскости вокруг ее точки на данный угол, параллельный перенос плоскости по самой себе в данном направлении. отражение относительно прямой, лежащей в плоскости, — определяются аналогичным образом и являются еще более наглядными, чем аналогичные преобразования пространства.

Взаимно однозначные преобразования. При рассмотрении всевозможных преобразований одного и того же множества прежде всего замечается фундаментальное различие между взаимно однозначными отображениями множества на себя и отображениями не взаимно однозначными. Преобразование A множества M называется взаимно однозначным отображением этого множества на себя, если не только каждому элементу множества M отвечает определенный единственный элемент множества M, — это содержится в определении преобразования, — но если также для каждого элемента y множества M существует один и только один элемент x, который переходит в элемент y. Иными словами, преобразование A является взаимно однозначным, если «уравнение» $xA=y$ для каждого y из M имеет одно и только одно «решение» x в M.

Все рассмотренные выше преобразования пространств — отражения, повороты и переносы — являются взаимно однозначными, так как при этом не только для каждой точки X существует точка, в которую X переходит, но также существует единственная точка, которая переходит в X.

Легко привести и противоположные примеры; так, преобразование множества чисел 1, 2, 3, 4, заданное таблицей $\begin{pmatrix}1&2&3&4\\2&1&2&3\end{pmatrix}$, не является взаимно однозначным, так как при нем в число 4 ничто не переходит. Преобразование множества всех натуральных чисел 1, 2, 3, ..., заданное таблицей $\begin{pmatrix}1&2&3&4&5&6&\cdots\\1&1&2&2&3&3&\cdots\end{pmatrix}$, также не является взаимно однозначным. Хотя здесь для каждого числа n имеется число $2n$, которое в него переходит, но число $2n$ не единственное, обладающее этим свойством, поскольку $2n-1$ также переходит в n. Вообще для преобразований, заданных таблицами, очень легко установить признак, при котором преобразования будут взаимно однозначными. Для этого, очевидно,

необходимо и достаточно, чтобы в нижней строке таблицы встречался каждый элемент множества и притом только один раз.

Иногда в математике рассматривают и не взаимно однозначные преобразования. Например, известно, какое большое значение имеет операция проектирования пространства на плоскость. Это преобразование не взаимно однозначное, так как при нем в каждую точку проектируется не одна, а целый ряд точек пространства. Но в большинстве случаев приходится иметь дело лишь с взаимно однозначными преобразованиями; эти преобразования, в частности, играют основную роль, когда рассматриваются физические процессы, при которых элементы изучаемой системы не сливаются друг с другом, не уничтожаются и не возникают.

В дальнейшем, говоря о преобразованиях, мы будем подразумевать преобразования взаимно однозначные; их часто называют также подстановками, особенно в случаях, когда речь идет о преобразованиях конечного множества.

Для каждого (взаимно однозначного) преобразования A множества M на себя легко определить обратное преобразование A^{-1}. Если преобразование A переводит произвольный элемент x множества M в элемент y, то преобразование, переводящее элемент y в x, называется обратным преобразованию A и обозначается A^{-1}. Например, если $A=\begin{pmatrix}1&2&3&4\\2&3&4&1\end{pmatrix}$, то $A^{-1}=\begin{pmatrix}2&3&4&1\\1&2&3&4\end{pmatrix}=\begin{pmatrix}1&2&3&4\\4&1&2&3\end{pmatrix}$; если A — поворот пространства вокруг оси на угол φ, то A^{-1} — поворот вокруг той же оси на угол φ в противоположном направлении, и т. д.

Иногда может случиться, что обратное преобразование будет совпадать с данным. Таким свойством, в частности, обладают отражения относительно плоскости или точки в пространстве. Тем же свойством обладает подстановка $A=\begin{pmatrix}2&1&4&3\\1&2&3&4\end{pmatrix}$, так как $A^{-1}=\begin{pmatrix}1&2&3&4\\2&1&4&3\end{pmatrix}=\begin{pmatrix}2&1&4&3\\1&2&3&4\end{pmatrix}$.

Заметим, что говорить об обратном преобразовании для преобразований не взаимно однозначных нельзя, так как эти преобразования в отдельные элементы могут или ничего не переводить или переводить несколько элементов.

Общее определение симметрии. В математике и ее приложениях очень редко возникает потребность рассмотрения всех преобразований данного множества. Дело в том, что сами множества редко приходится мыслить только как простое объединение своих элсментов, ничем не связанных друг с другом. Это и естественно, так как множества, рассматриваемые в математике, являются отвлеченными образами реальных совокупностей, элементы которых всегда находятся в бесконечном числе взаимосвязей друг с другом и в связях с тем, что находится за пределами рассматриваемого множества. При этом в математике приходится отвлекаться от большей части этих связей, но наиболее суще-

ственные сохранять и учитывать. Это заставляет в первую очередь рассматривать такие преобразования множеств, которые не нарушают тех или иных учитываемых связей между элементами. Такие преобразования часто называют допустимыми преобразованиями или *автоморфизмами* по отношению к учитываемым связям элементов множества. Например, для точек пространства важным является понятие расстояния между двумя точками. Наличие этого понятия учитывает связь между точками, заключающуюся в том, что любые две точки находятся на определенном расстоянии одна от другой. Преобразованиями, не нарушающими этих связей, являются такие преобразования, при которых расстояния между точками не изменяются. Эти преобразования называются «*движениями*» пространства.

Пользуясь понятием автоморфизма, не трудно дать общее определение симметрии. Пусть дано некоторое множество M, в котором учитываются определенные связи между элементами, и пусть P есть некоторая часть M. Говорят, что совокупность P симметрична или инвариантна относительно допустимого преобразования A множества M, если преобразование A переводит каждый элемент множества P снова в элемент множества P. Поэтому симметрия множества P характеризуется совокупностью допустимых преобразований объемлющего множества M, преобразующих P в себя. Понятие симметрии тел в пространстве вполне подходит под данное определение. Роль множества M играет все пространство, роль допустимых преобразований — «движения», роль P — данное тело. Симметрия тела P характеризуется, таким образом, совокупностью движений, при которых тело P совмещается с собой.

Рассмотренные ранее отражения, параллельные переносы и повороты пространства около заданной прямой являются частными случаями движений, так как расстояния между точками при этих преобразованиях, очевидно, не меняются. Более подробное исследование показывает, что каждое движение плоскости есть либо перенос, либо поворот около центра, либо отражение относительно прямой, либо комбинация отражения относительно прямой с переносом вдоль этой прямой. Аналогично каждое движение пространства есть либо параллельный перенос, либо поворот около оси, либо винтовое движение, т. е. поворот вокруг оси, сопровождаемый переносом вдоль этой оси, либо же отражение относительно плоскости, сопровождаемое, может быть, еще переносом вдоль плоскости отражения или поворотом вокруг перпендикулярной к этой плоскости оси.

Параллельные переносы, повороты и винтовые движения пространства называются его собственными движениями, или движениями 1-го рода. Остальные «движения» (включающие в себя отражение) носят название несобственных движений, или движений 2-го рода. На плоскости движениями 1-го рода будут параллельные переносы и повороты,

а отражения относительно прямой и отражения, сопровождаемые поворотом или переносом, будут движениями 2-го рода.

Легко сообразить, что преобразования, являющиеся движениями 1-го рода, можно получить как результат непрерывного движения пространства самого в себе или плоскости самой по себе. Движения 2-го рода таким образом получить невозможно, так как этому препятствуют зеркальные отражения, входящие в их состав.

Часто говорят, что плоскость симметрична во всех своих частях или что все точки плоскости равноправны. На точном языке преобразований это утверждение означает, что любую точку плоскости можно совместить с любой другой ее точкой посредством подходящего «движения».

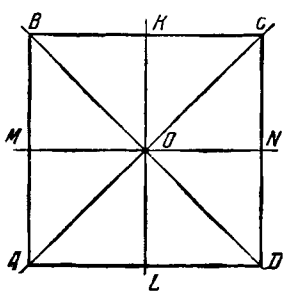


Рис. 7.

Рассмотренные ранее случаи симметрии тел или фигур также охватываются общим определением симметрии. Так, например, тело, симметричное относительно плоскости α, совмещается само с собой при отражении относительно плоскости α; тело, симметричное относительно центра O, совмещается само с собой при отражении относительно O. Поэтому степень симметричности тела или пространственной фигуры вполне характеризуется совокупностью тех движений 1-го и 2-го рода пространства, которые совмещают тело или фигуру самое с собой. Чем богаче и разнообразнее указанная совокупность движений, тем большей степенью симметричности обладает тело или фигура. В частности, если эта совокупность не содержит никаких движений, кроме тождественного преобразования, то тело можно назвать несимметричным.

Степень симметричности квадрата на плоскости характеризуется совокупностью движений плоскости, совмещающих квадрат сам с собой. Но если квадрат совмещается сам с собой, то точка пересечения его диагоналей также должна совмещаться сама с собой. Поэтому искомые движения оставляют центр квадрата неподвижным и потому являются либо поворотами около центра, либо отражениями относительно прямых, проходящих через центр. Из рис. 7 легко усматриваем, что квадрат $ABCD$ симметричен относительно поворотов вокруг его центра O на углы, кратные 90°, а также по отношению к отражениям относительно диагоналей AC, BD и прямых KL, MN. Эти восемь движений и характеризуют симметрию квадрата.

Совокупность симметрий прямоугольника сводится к поворотам около центра на 180° и отражениям относительно прямых, соединяющих середины противоположных сторон, а совокупность симметрий параллелограмма (рис. 8) состоит лишь из поворотов вокруг центра на углы, кратные 180°, т. е. из отражения относительно центра и тождественного преобразования.

Выше мы приводили алгебраический пример симметрии; именно, было отмечено, что имеет смысл понятие симметрии многочлена от нескольких переменных.

Рассмотрим, как можно охарактеризовать симметрию многочлена. Будем говорить, что в многочлене $F(x_1, x_2, \ldots, x_n)$ сделана подстановка неизвестных $A = \begin{pmatrix} x_1, & x_2, & \ldots, & x_n \\ x_{i_1}, & x_{i_2}, & \ldots, & x_{i_n} \end{pmatrix}$ или короче $A = \begin{pmatrix} 1, & 2, & \ldots, & n \\ i_1, & i_2, & \ldots, & i_n \end{pmatrix}$, если в данный многочлен всюду вместо буквы x_1 поставлена буква x_{i_1}, вместо x_2 поставлена x_{i_2} и т. д. Полученный таким образом многочлен обозначается через FA. Так, если $F = x_1^2 - 2x_2 + x_3 - x_4$, $A = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 3 & 1 & 4 & 2 \end{pmatrix}$, то $FA = x_3^2 - 2x_1 + x_4 - x_2$.

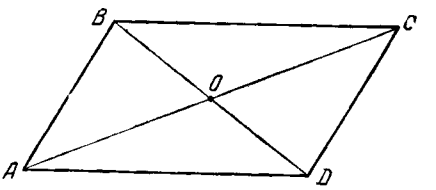


Рис. 8.

Симметрия данного многочлена характеризуется совокупностью тех подстановок неизвестных, которые, будучи выполнены над многочленом, его не изменяют. Например, симметрия многочлена $x_1^3 + 2x_2 + x_3^3 + 2x_4$ характеризуется четырьмя подстановками: $\begin{pmatrix} 1 & 2 & 3 & 4 \\ 1 & 2 & 3 & 4 \end{pmatrix}$, $\begin{pmatrix} 1 & 2 & 3 & 4 \\ 3 & 2 & 1 & 4 \end{pmatrix}$, $\begin{pmatrix} 1 & 2 & 3 & 4 \\ 1 & 4 & 3 & 2 \end{pmatrix}$, $\begin{pmatrix} 1 & 2 & 3 & 4 \\ 3 & 4 & 1 & 2 \end{pmatrix}$, а симметрия многочлена $x_1^3 + 2x_2 + x_3^3 + x_4$ характеризуется лишь двумя подстановками: $\begin{pmatrix} 1 & 2 & 3 & 4 \\ 1 & 2 & 3 & 4 \end{pmatrix}$ и $\begin{pmatrix} 1 & 2 & 3 & 4 \\ 3 & 2 & 1 & 4 \end{pmatrix}$.

§ 3. ГРУППЫ ПРЕОБРАЗОВАНИЙ

Умножение преобразований. Изучая свойства преобразований, легко заметить, что некоторые преобразования можно составить из нескольких других. Так, винтовые движения составляются из поворотов вокруг оси и сдвигов вдоль оси. Этот процесс составления новых преобразований из заданных носит название умножения преобразований. Производя над произвольным элементом x множества M какое-либо преобразование A и затем к новому элементу xA применяя преобразование B, получим элемент $(xA)B$. Преобразование, переводящее x непосредственно в $(xA)B$, называется произведением преобразования A на B и обозначается через AB. Следовательно, по определению, имеем

$$x(AB) = (xA)B.$$

Пример:

$$\begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 3 & 4 & 1 \end{pmatrix}\begin{pmatrix} 1 & 2 & 3 & 4 \\ 3 & 4 & 1 & 2 \end{pmatrix} = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 4 & 1 & 2 & 3 \end{pmatrix}.$$

Действительно, первая подстановка переводит 1 в 2, а вторая 2 переводит в 4, поэтому результирующая подстановка должна переводить 1 в 4 и т. д. Вот еще несколько примеров:

$$\begin{pmatrix}1&2&3&4\\3&1&4&2\end{pmatrix}\begin{pmatrix}1&2&3&4\\1&3&2&4\end{pmatrix}^{-1}=\begin{pmatrix}1&2&3&4\\3&1&4&2\end{pmatrix}\begin{pmatrix}1&2&3&4\\1&3&2&4\end{pmatrix}=\begin{pmatrix}1&2&3&4\\2&1&4&3\end{pmatrix};$$

$$\begin{pmatrix}1&2&3&4\\2&1&4&3\end{pmatrix}\begin{pmatrix}1&2&3&4\\3&1&2&4\end{pmatrix}=\begin{pmatrix}1&2&3&4\\1&3&4&2\end{pmatrix};\quad\begin{pmatrix}1&2&3&4\\3&1&2&4\end{pmatrix}\begin{pmatrix}1&2&3&4\\2&1&4&3\end{pmatrix}=\begin{pmatrix}1&2&3&4\\4&2&1&3\end{pmatrix}.$$

Последние два примера показывают, что умножение преобразований есть, как говорят, действие некоммутативное: его результат зависит от порядка сомножителей. Это же легко подтверждается и для умножения движений плоскости. Пусть, например, A есть поворот плоскости на 90° вокруг начала O, а B — перенос вдоль оси Ox на единицу.

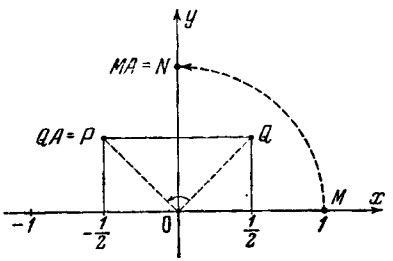


Рис. 9.

Посмотрим, во что переводят преобразования AB и BA точку O. По определению, имеем (рис. 9)

$$O(AB)=(OA)B=OB=M,$$

$$O(BA)=(OB)A=MA=N,$$

т. е. $AB\neq BA$.

Чтобы ближе выяснить геометрическую природу преобразования BA, рассмотрим точку P. Имеем

$$P(BA)=(PB)A=QA=P,$$

т. е. точка P остается при преобразовании BA неподвижной. Исходя из этого, легко показать, что BA есть просто поворот плоскости на 90° вокруг точки P. Аналогично

$$Q(AB)=(QA)B=PB=Q,$$

и AB есть поворот плоскости на 90° вокруг точки Q.

Умножение движений плоскости или пространства происходит в общем случае по довольно сложным законам. Однако в двух важных случаях законы умножения очень просты. Во-первых, если умножаются повороты плоскости вокруг одной и той же точки или повороты пространства вокруг одной и той же прямой на углы φ и ψ, то результирующее преобразование будет соответственным поворотом на угол $\varphi+\psi$. Во-вторых, если умножаются переносы, характеризуемые векто-

рами $\overrightarrow{MN}$ и $\overrightarrow{NP}$, то произведение будет также переносом, характеризуемым вектором $\overrightarrow{MP}$, т. е. суммой первоначальных векторов.

Сам термин «умножение» преобразований вызван некоторой аналогией между умножением чисел и умножением преобразований. Однако эта аналогия неполная. Так, при умножении чисел имеет место коммутативный (переместительный) закон. Мы уже видели, что при умножении преобразований этот закон может быть нарушен. Второй же основной закон арифметики — сочетательный закон (ассоциативность) умножения — в полной мере сохраняется для преобразований. Именно, для любых преобразований A, B, C множества M имеет место равенство $A(BC)=(AB)C$.

В самом деле, если m — произвольный элемент из M, то

$$m[A(BC)]=(mA)(BC)=[(mA)B]C=[m(AB)]C=m[(AB)C].$$

Ассоциативный закон позволяет вместо двух произведений $A(BC)$ и $(AB)C$ преобразований A, B, C говорить лишь об одном произведении $A(BC)=(AB)C=ABC$. Этот же закон показывает, что и произведение четырех и более преобразований не зависит от расстановки скобок.

Далее, среди преобразований имеется преобразование, играющее роль числа 1, это — тождественное или единичное преобразование E, которое оставляет каждый элемент множества M неизменным. Ясно, что $AE=EA=A$, каково бы ни было преобразование A.

Отметим еще следующий важный факт: произведение взаимно однозначных преобразований есть преобразование взаимно однозначное. В самом деле, чтобы найти элемент x множества M, который произведением AB переводится в данный элемент a, достаточно найти элемент x_1, переводимый преобразованием B в a, и затем найти элемент x_2, переводимый преобразованием A в x_1. Так как $x_2(AB)=(x_2A)B=x_1A=a$, то x_2 и будет искомым элементом x.

Произведение преобразования A на обратное преобразование A^{-1} есть единичное преобразование, т. е.

$$AA^{-1}=A^{-1}A=E.$$

Это непосредственно следует из определения обратного преобразования.

Рассмотренный выше пример умножения переноса плоскости на поворот показывает, что свойства произведения преобразований не всегда легко усмотреть, исходя из свойств сомножителей. Однако произведение преобразований вида $C=B^{-1}AB$ представляет важное исключение: свойства C здесь очень просто связаны со свойствами A и B. Именно, если элемент m множества M преобразованием A переводится в n, то «сдвинутый» посредством преобразования B элемент mB преобразованием C переводится в «сдвинутый» элемент nB.

Доказательство: $(mB)B^{-1}AB=mAB=nB$.

Преобразование $B^{-1}AB$ называют полученным из A путем преобразования его при помощи B или *сопряженным* с A посредством B.

Преобразуем, например, поворот P_O плоскости около точки O при помощи переноса V. Согласно приведенному правилу, чтобы найти пары начальных и конечных положений точек для преобразованного движения $C=V^{-1}P_OV$, надо сдвинуть при помощи V соответственные пары точек для преобразования P_O. Так как точка O при повороте P_O отвечает самой себе (рис. 10), то и точка OV относительно преобразования C отвечает самой себе. Далее, если точка M переводится преобразованием P в точку N, то сдвинутая точка MV будет переводиться преобразованием C в точку NV. Из рис. 10, таким образом, видно, что преобразование C будет поворотом около точки OV на тот же угол φ, что и поворот P.

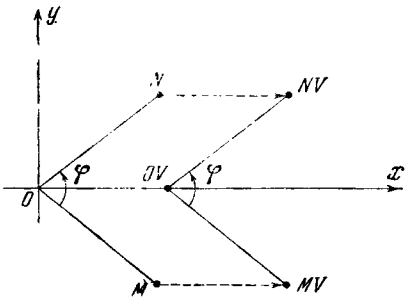


Рис. 10.

Аналогично может быть доказано, что если перенос плоскости, характеризуемый вектором $\overrightarrow{MN}$, преобразовать посредством поворота P_O на угол φ, то получится снова перенос плоскости, характеризуемый повернутым вектором.

Указанное выше правило для отыскания преобразования $B^{-1}AB$ весьма изящно формулируется также в случае, когда преобразования задаются таблицами. Пусть

$$A=\begin{pmatrix}1\,2\ldots n\\ a_1a_2\ldots a_n\end{pmatrix},\quad B=\begin{pmatrix}1\,2\ldots n\\ b_1b_2\ldots b_n\end{pmatrix},$$

тогда

$$B^{-1}AB=\begin{pmatrix}b_1b_2\ldots b_n\\ 1\,2\ldots n\end{pmatrix}\begin{pmatrix}1\,2\ldots n\\ a_1a_2\ldots a_n\end{pmatrix}\begin{pmatrix}1\,2\ldots n\\ b_1b_2\ldots b_n\end{pmatrix}=\begin{pmatrix}b_1\,b_2\ldots b_n\\ b_{a_1}b_{a_2}\ldots b_{a_n}\end{pmatrix},$$

т. е. чтобы преобразовать подстановку A при помощи подстановки B, нужно все элементы верхней и нижней строк подстановки A подвергнуть преобразованию, предусмотренному подстановкой B. Например, если

$$A=\begin{pmatrix}1\,2\,3\,4\,5\\ 3\,5\,4\,1\,2\end{pmatrix},\quad B=\begin{pmatrix}1\,2\,3\,4\,5\\ 2\,5\,1\,3\,4\end{pmatrix},$$

то

$$B^{-1}AB=\begin{pmatrix}1B\;2B\;3B\;4B\;5B\\ 3B\;5B\;4B\;1B\;2B\end{pmatrix}=\begin{pmatrix}2\,5\,1\,3\,4\\ 1\,4\,3\,2\,5\end{pmatrix}=\begin{pmatrix}1\,2\,3\,4\,5\\ 3\,1\,2\,5\,4\end{pmatrix}.$$

Заметим, что хотя в общем случае произведение двух преобразований зависит от порядка сомножителей, в отдельных случаях произ-

ведения AB и BA могут быть одинаковыми. Тогда преобразования A и B называются перестановочными или коммутирующими. Если $AB = BA$, то

$$B^{-1}AB = B^{-1}BA = A.$$

Таким образом, преобразование данной подстановки при помощи коммутирующей с ней подстановки не меняет данную подстановку.

Группы преобразований. Совокупности преобразований, характеризующих симметрию некоторой фигуры, не могут быть произвольными, они заведомо должны обладать следующими свойствами:

1. Произведение двух преобразований, принадлежащих совокупности, само принадлежит этой совокупности.

2. Тождественное преобразование принадлежит совокупности.

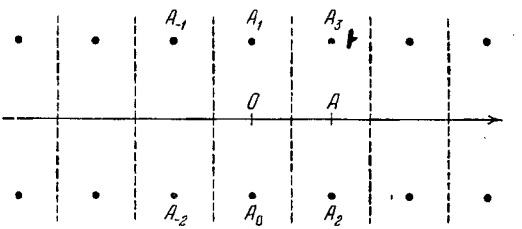


Рис. 11.

3. Если преобразование принадлежит совокупности, то обратное преобразование также принадлежит совокупности.

Эти свойства оказались очень важными для изучения преобразований, ввиду чего всякую совокупность взаимно однозначных преобразований множества M, обладающую перечисленными тремя свойствами, стали называть *группой преобразований* множества M, независимо от того, характеризует эта совокупность симметрию некоторой фигуры или нет.

С точки зрения алгебры свойства 1—3 являются очень существенными, так как они позволяют, исходя из некоторых преобразований A, B, C,..., принадлежащих заданной совокупности, составлять различные новые преобразования вида $ABAC$, $A^{-1}BCB^{-1}$ и т. п., причем свойства 1—3 гарантируют, что все получаемые преобразования не выходят за пределы заданной совокупности преобразований.

Число преобразований, составляющих группу, называется *порядком группы*; последний может быть и конечным и бесконечным. Соответственно этому и группы разделяются на конечные и бесконечные. Выше была рассмотрена группа симметрий квадрата на плоскости. Эта группа оказалась содержащей всего восемь преобразований. С другой стороны, бесконечная совокупность точек A_i плоскости, изображенная на рис. 11, преобразуется в себя следующими движениями плоскости: переносами вдоль оси OA в том и другом направлениях на расстояния, кратные расстоянию OA; отражениями относительно пунктирных прямых; отражением относительно оси OA. Отсюда видно, что группа симметрий этой фигуры бесконечная.

Совокупность преобразований, сохраняющих некоторый объект, т. е. характеризующих его симметрию, всегда является группой. Этот

способ задания групп в виде групп симметрий принадлежит к числу важнейших. По этому принципу получаются наиболее важные группы. К ним прежде всего следует отнести группы движений плоскости и пространства. Большой интерес представляют также группы симметрий правильных многогранников. Как известно, в пространстве существует всего пять типов правильных многогранников (4, 6, 8, 12 и 20-гранники). Беря какой-либо из правильных многогранников и рассматривая все движения пространства, совмещающие данный многогранник с собой, мы получим группу — группу симметрий этого многогранника. Если вместо всех движений рассматривать только движения 1-го рода, совмещающие многогранник с самим собой, то получим опять группу, являющуюся частью полной группы симметрий многогранника. Эта группа называется группой вращений многогранника. Поскольку при совмещении многогранника с собой его центр также совмещается с самим собой, то все движения, входящие в группу симметрий многогранника, оставляют неподвижным центр многогранника и потому могут быть только или поворотами около осей, проходящих через центр, или отражениями относительно плоскостей, проходящих через центр, или, наконец, отражениями в таких плоскостях, сопровождаемыми поворотами вокруг осей, проходящих через центр и перпендикулярных к этим плоскостям.

Пользуясь этим замечанием, легко найти все группы симметрий и группы вращений правильных многогранников. В табл. 1 указаны порядки групп симметрий и групп вращений правильных многогранников. Все эти группы являются конечными.

Таблица 1

Число граней	4	6	8	12	20
Порядок группы симметрий . .	24	48	48	120	120
Порядок группы вращений . .	12	24	24	60	60

Группы подстановок. Из групп преобразований исторически первыми рассматривались в математике группы подстановок переменных $x_1, x_2, \ldots, x_n$ в многочленах от этих переменных. Рассмотрение таких групп тесно связано с вопросом о решении в радикалах уравнений высших степеней. Очевидно, что совокупность всех подстановок переменных, не меняющих значений одного или нескольких многочленов от этих переменных, является группой. Многочлены, не меняющиеся при всех подстановках переменных, называются симметрическими многочленами. Например, $x_1 + x_2 + \ldots + x_n$ есть симметрический многочлен. Соответственно, совокупность всех подстановок дан-

ного множества переменных называется *симметрической* группой подстановок этого множества.

Число переставляемых переменных называется степенью симметрической группы. Вместо подстановок переменных $x_1, \ldots, x_n$ можно рассматривать просто подстановки чисел 1, 2, ..., n. Так как каждую подстановку чисел можно записать в виде $\begin{pmatrix} 1 & 2 & \ldots & n \\ a_1 & a_2 & \ldots & a_n \end{pmatrix}$, где $a_1, a_2, \ldots, a_n$ — числа 1, 2, ..., n, записанные в некотором порядке, то число всех подстановок равно числу перестановок n элементов, т. е. порядок симметрической группы равен $n! = 1 \cdot 2 \cdot 3 \ldots n$. Этот порядок очень быстро растет со степенью n, и порядок группы подстановок 10 переменных равен уже 3 628 800.

Рассмотрим многочлен

$$F(x_1, \ldots, x_n) = \\ = (x_2 - x_1)(x_3 - x_1)\ldots(x_n - x_1)(x_3 - x_2)\ldots(x_n - x_2)\ldots(x_n - x_{n-1}). \quad (1)$$

Ясно, что каждая подстановка переменных или оставляет величину многочлена F неизменной или меняет лишь его знак. Подстановки первого рода называются четными. Подстановки, меняющие знак F, называются нечетными. Совокупность четных подстановок образует группу симметрий многочлена (1). Она называется знакопеременной группой подстановок.

Произведение двух четных подстановок есть четная подстановка, ибо четные подстановки образуют группу. Произведение двух нечетных подстановок есть подстановка четная.

Действительно, если A и B — нечетные подстановки, то

$$FAB = (FA)B = (-F)B = -(-F) = F.$$

Таким же образом доказывается, что произведение четной и нечетной подстановок есть нечетная подстановка и что подстановка обратная к четной или нечетной подстановке есть подстановка той же четности.

Примером нечетной подстановки может служить подстановка $S = \begin{pmatrix} 1, & 2, & 3, \ldots, & n \\ 2, & 1, & 3, \ldots, & n \end{pmatrix}$, меняющая местами элементы 1 и 2.

Разложение подстановок на циклы. При изучении групп подстановок большое значение имеет представление подстановок в виде произведений так называемых циклов. По определению символом $(m_1, m_2, \ldots, m_k)$ обозначается подстановка, переводящая m_1 в m_2, m_2 в $m_3, \ldots, m_{k-1}$ в m_k, m_k снова в m_1, а все остальные элементы рассматриваемого множества оставляющая на месте. Так, например, если рассматриваются подстановки чисел 1, 2, 3, 4, 5, то

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

Подстановка вида $(m_1, m_2, \ldots, m_k)$ называется *циклической* или *циклом* длины k; $m_1, m_2, \ldots, m_k$ называются элементами цикла. Единичная подстановка условно записывается в виде циклов $(1)=(2)=\ldots$ длины 1. Циклы длины 2 называются *транспозициями*. Переставляя элементы цикла в циклическом порядке, мы получаем ту же самую подстановку, например $(1, 2, 3)=(2, 3, 1)=(3, 1, 2)$, $(5, 6)=(6, 5)$.

Легко проверяется, что циклы без общих элементов, например $(2, 3)$ и $(1, 4, 5)$, являются коммутирующими подстановками, и поэтому при перемножении таких циклов можно не обращать внимания на порядок сомножителей в произведении.

Значение циклов в общей теории основывается на следующей теореме: всякая подстановка может быть представлена в виде произведения циклов без общих элементов, причем это представление однозначно с точностью до порядка сомножителей.

Доказательство теоремы непосредственно видно из способа такого представления. Допустим, что мы желаем разложить подстановку $A=\begin{pmatrix}1&2&3&4&5&6\\4&5&6&3&2&1\end{pmatrix}$. Мы видим, что A переводит 1 в 4, 4 в 3, 3 в 6, 6 в 1. В результате имеем первый множитель $(1, 4, 3, 6)$. Берем оставшееся число 2 и видим, что A переводит 2 в 5, 5 в 2. Поэтому второй множитель будет $(2, 5)$. Поскольку все числа рассмотрены, то

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

Разложение подстановок на циклы, имеющие общие элементы, также возможно, но уже неоднозначно. Например,

$$(a_1, a_2, \ldots, a_n)=(a_1, a_2)(a_1, a_3)\ldots(a_1, a_n)=(a_2, a_3)(a_2, a_4)\ldots(a_2, a_n)(a_2, a_1). \tag{3}$$

Докажем, что каждый двойной цикл есть нечетная подстановка. Мы уже убедились в этом для цикла $(1, 2)$. Но любой цикл (i, j) есть $S^{-1}(1, 2)S$, где S — любая подстановка $\begin{pmatrix}1,&2,&\ldots\\i,&j,&\ldots\end{pmatrix}$, переводящая 1 в i и 2 в j. Подстановка $S^{-1}(1, 2)S$ есть нечетная подстановка, ибо $(1, 2)$ нечетная, а S и S^{-1} одновременно или четные или нечетные.

Согласно формуле (3), цикл длины $m+1$ может быть представлен в виде произведения m нечетных подстановок. Поэтому цикл длины $m+1$ есть подстановка нечетная, если $m+1$ четно, и четная, если $m+1$ нечетно. Это дает возможность быстро подсчитать четность подстановок, разложение которых на циклы известно. В частности, подстановка $\begin{pmatrix}1&2&3&4&5&6\\4&5&6&3&2&1\end{pmatrix}$ четна, так как она, согласно формуле (2), есть произведение двух нечетных подстановок.

Подгруппы. Часть группы, сама являющаяся группой, называется *подгруппой* данной группы. Так, знакопеременная группа подстановок

переменных $x_1, x_2, \ldots, x_n$ является подгруппой симметрической группы подстановок этих переменных. Совокупность собственных движений плоскости есть группа, являющаяся подгруппой группы всех собственных и несобственных движений плоскости.

С формальной точки зрения единичное (тождественное) преобразование само по себе уже образует подгруппу. Точно так же и любая группа может быть рассматриваема как своя подгруппа. Однако, кроме этих тривиальных подгрупп, почти всегда группы содержат много других подгрупп. Знание всех подгрупп данной группы дает довольно полное представление о внутреннем строении заданной группы.

Одним из распространенных способов образования подгрупп является указание так называемых образующих подгрупп.

Пусть $A_1, A_2, \ldots, A_m$ — какие-либо преобразования, принадлежащие группе G. Совокупность H всех преобразований, которые можно получить, перемножая между собой в произвольном числе заданные преобразования и преобразования, им обратные, будет группой. Действительно, единичное преобразование принадлежит к этой совокупности, так как его можно представить в виде $A_1A_1^{-1}$. Далее, если преобразования B и C можно представить в виде произведений, то, перемножая эти произведения, мы получим требуемое представление и для BC. Наконец, если B выражается в виде произведения, например, $B = A_1^{-1}A_2A_1A_1A_2^{-1}$, то и B^{-1} можно представить в виде требуемого произведения, так как $B^{-1} = A_2A_1^{-1}A_1^{-1}A_2^{-1}A_1$.

Группа H является, очевидно, подгруппой группы G и называется подгруппой, порожденной преобразованиями $A_1, \ldots, A_m$, а сами преобразования $A_1, \ldots, A_m$ называются образующими подгруппы H. Может случиться, что H совпадает с G, тогда $A_1, \ldots, A_m$ называются образующими самой группы G. Нетрудно убедиться на примерах, что одна и та же подгруппа может порождаться самыми различными системами образующих.

Подгруппа, порождаемая одним преобразованием A, называется циклической подгруппой. Ее элементами являются преобразования

$$E, A, AA, AAA, \ldots, A^{-1}, A^{-1}A^{-1}, A^{-1}A^{-1}A^{-1}, \ldots,$$

которые, естественно, называются степенями преобразования A. Именно:

$$E = A^0, \quad A = A^1, \quad AA = A^2, \ldots, \quad A^{-1}A^{-1} = A^{-2}, \quad A^{-1}A^{-1}A^{-1} = A^{-3}, \ldots$$

Легко доказывается, как в обычной арифметике, что

$$A^mA^n = A^{m+n} \text{ и } (A^m)^n = A^{mn}. \tag{4}$$

Преобразование называется *периодическим*, если некоторая его положительная степень равна тождественному преобразованию. Наименьший положительный показатель степени, в которую нужно возвести периодическое преобразование, чтобы получить тождественное, назы-

вается *порядком преобразования*. Условно говорят, что порядок непериодического преобразования бесконечен.

Рассмотрим несколько примеров. Пусть A — поворот плоскости вокруг точки O на $\frac{360^\circ}{n}$, где n — данное положительное целое число, большее единицы. Тогда A^2 будет поворотом на угол $2\frac{360^\circ}{n}$, A^3 — поворотом на $3\frac{360^\circ}{n}$, A^{n-1} — поворотом на $(n-1)\frac{360^\circ}{n}$, A^n — поворотом на 360°, т. е. будет тождественным преобразованием. Это показывает, что поворот на $\frac{360^\circ}{n}$ есть периодическое преобразование порядка n.

Пусть A — перенос плоскости вдоль некоторой прямой. Тогда A^2, A^3, ... будут также переносами вдоль той же прямой соответственно на двойное, тройное и т. д. расстояния. Поэтому никакая положительная степень A не есть тождественное преобразование, и порядок A бесконечен.

Элементами циклической группы, порожденной элементом A, являются

$$\ldots, A^{-2}, A^{-1}, E, A, A^2, \ldots \tag{5}$$

Если A — преобразование бесконечного порядка, то все преобразования в последовательности (5) различны, и группа бесконечна. Действительно, в противном случае мы имели бы равенство вида $A^k = A^l$ $(k < l)$, откуда $A^{l-k} = E$, $l-k > 0$, что противоречит непериодичности A.

Предположим теперь, что A — периодическое преобразование порядка m. Тогда

$$A^m = E,\ A^{m+1} = A,\ A^{m+2} = A^2,\ \ldots,\ A^{m-1} = A^{-1},\ A^{m-2} = A^{-2},\ \ldots,$$

т. е. последовательность (5) состоит из одних и тех же периодически повторяющихся преобразований E, A, A^2, ..., A^{m-1}. Они между собою различны, так как, если бы оказалось $A^k = A^l$ $(0 \leqslant k < l < m)$, то было бы $A^{l-k} = E$, $0 < l-k < m$, что противоречит выбору m. Следовательно, циклическая подгруппа, порожденная преобразованием порядка m, содержит ровно m различных преобразований.

Группа, все элементы которой коммутируют друг с другом, называется коммутативной или *абелевой*, по имени норвежского математика Абеля, открывшего важное значение таких групп для теории уравнений.

Формулы (4) показывают, что степени одного и того же преобразования всегда коммутируют друг с другом: $A^m A^n = A^n A^m = A^{m+n}$. Поэтому циклические подгруппы всегда абелевы.

В арифметике чисел наряду с умножением важную роль играет действие деления. В теории групп, вследствие некоммутативности умно-

жения, приходится говорить о двух делениях: правом и левом. В самом деле, решение уравнения $Ax=B$, где A, B — данные преобразования, а x — искомое преобразование, естественно назвать правым частным, а решение уравнения $yA=B$ — левым частным от деления B на A. Умножая обе части первого равенства на A^{-1} слева, а обе части второго на A^{-1} справа, получим: $x=A^{-1}B$, $y=BA^{-1}$. Таким образом, в качестве «частного» преобразований B и A можно рассматривать $A^{-1}B$ или BA^{-1}.

На многочисленных примерах мы видели, что в группах, вообще говоря, $AB\neq BA$. «Частное» $(AB)(BA)^{-1}$ или $(BA)^{-1}(AB)$ можно принять за «меру» некоммутируемости подстановок A и B. Второе из этих выражений, именно $(BA)^{-1}(AB)=A^{-1}B^{-1}AB$, называют *коммутатором* A и B и обозначают (A, B). Из формулы

$$(A, B)=A^{-1}B^{-1}AB$$

следует, что коммутатор можно представлять себе и как частное от «деления» сопряженного преобразования $B^{-1}AB$ на первоначальное A.

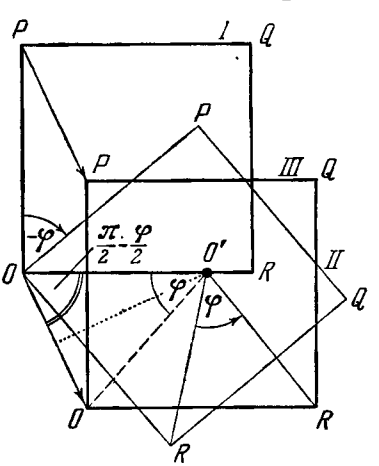


Рис. 12.

Если, например, A — перенос плоскости, то сопряженное преобразование будет также переносом, а частное двух переносов, очевидно, есть перенос. Поэтому коммутатор переноса и любого движения плоскости есть перенос. Пусть A — поворот на угол φ вокруг некоторой точки O, а B — поворот или перенос. Тогда сопряженное преобразование будет снова поворотом на угол φ, но вокруг смещенной точки O'. Следовательно, коммутатор (A, B) в рассматриваемом случае есть произведение поворота вокруг точки O на отрицательный угол φ и поворота вокруг точки O' на положительный угол φ. Из рис. 12 видно, что результирующее преобразование есть перенос под углом $\frac{\pi}{2}-\frac{\varphi}{2}$ к отрезку $\overline{OO'}$ на расстояние $2\cdot\overline{OO'}\sin\frac{\varphi}{2}$.

Таким образом, мы приходим к интересному факту, что для плоскости коммутатор любых двух движений 1-го рода есть параллельный перенос или тождественное преобразование. Поскольку $(A, B)=E$ означает, что $AB=BA$, то на плоскости любая некоммутативная группа движений 1-го рода содержит параллельные переносы.

Подгруппа, порожденная коммутаторами всевозможных элементов группы G, называется *коммутантом* группы G. Вспоминая соответствующие определения, можно сказать, что коммутант группы G состоит из тех и только тех ее элементов, которые можно представить в виде

произведения коммутаторов. Поскольку для плоскости коммутатор любых двух движений 1-го рода есть параллельный перенос, а произведения параллельных переносов суть также параллельные переносы, можно утверждать, что коммутант группы движений плоскости 1-го рода состоит лишь из параллельных переносов.

Коммутант абелевой группы состоит из единичного преобразования, так как из $AB = BA$ следует $(A, B) = E$.

Пусть G — симметрическая группа всех подстановок чисел $1, 2, \ldots, n$. Покажем, что коммутатор любых двух подстановок A, B всегда будет подстановкой четной. В самом деле, подстановки AB, BA, а следовательно и $(BA)^{-1}$, всегда имеют одинаковую четность; тогда коммутатор $(A, B) = (BA)^{-1}(AB)$, как произведение подстановок одинаковой четности, есть подстановка четная.

Мы видим, что коммутант симметрической группы состоит лишь из четных подстановок. Легко можно доказать, что он совпадает со всей знакопеременной группой.

Коммутант группы G называется также производной группой и обозначается через G'; коммутант коммутанта G называется вторым коммутантом группы G и обозначается через G''. Повторяя этот процесс, мы получим определение коммутанта любого порядка группы G.

Если из коммутантов группы G хотя бы один (а тогда и все последующие) состоит лишь из единичного преобразования, то группа G называется *разрешимой*. Название это возникло в теории уравнений, где разрешимости группы соответствует разрешимость уравнения в радикалах. Группа движений 1-го рода для плоскости разрешима, так как уже ее 2-й коммутант равен единице. Симметрические группы 2, 3 и 4-й степени также разрешимы, поскольку их соответственно 1, 2 и 3-й коммутанты обращаются в единицу. Напротив, симметрические группы 5-й и более высоких степеней неразрешимы, так как можно доказать, что их 2-й коммутант совпадает с 1-м и отличен от единицы.

§ 4. ФЕДОРОВСКИЕ ГРУППЫ

Группы симметрий конечных плоских фигур. Как уже было установлено, симметричность фигуры или тела характеризуется группой движений плоскости или пространства, совмещающих фигуру с собой.

Наиболее просто находятся группы симметрий конечных фигур на плоскости[1]. В самом деле, пусть дана какая-либо конечная фигура на плоскости и пусть эта фигура совмещается сама с собой некоторым движением A. Тогда центр тяжести фигуры O должен движением A

[1] Конечность понимается в том смысле, что вся фигура расположена в ограниченной части плоскости, например в некотором круге.

также совмещаться сам с собой, т. е. A есть или поворот около O, или отражение относительно прямой, проходящей через O. Итак, группа симметрий любой конечной плоской фигуры может состоять лишь из поворотов около ее центра тяжести и из отражений относительно прямых, проходящих через центр.

Рассмотрим последовательно различные случаи, которые могут представиться при рассмотрении групп симметрий конечной плоской фигуры.

1. Группа симметрий K_1 состоит лишь из единичного (тождественного) преобразования. Это — группа симметрий любой несимметричной фигуры (рис. 13).

2. Группа симметрий K_2 состоит из единичного преобразования и отражения около одной прямой (рис. 14).

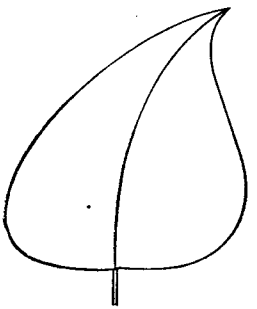


Рис. 13.

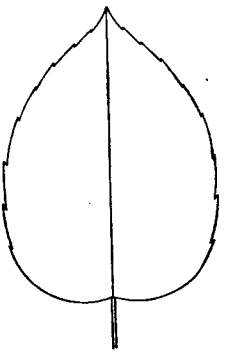


Рис. 14.

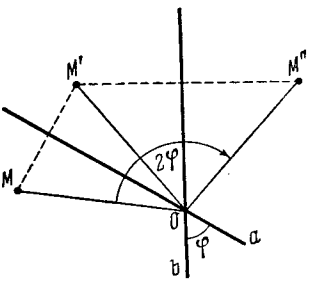


Рис. 15.

Заметим, что если группа K содержит отражения около двух прямых a, b, проходящих через O и образующих угол φ между собой, то произведение этих отражений будет поворотом вокруг O на угол 2φ (рис. 15). Отсюда видно, что группа K_2 — единственная из групп симметрий, не содержащая поворотов.

3. Группа симметрий K_3 состоит лишь из одних поворотов, среди которых нет поворотов на сколь угодно малый угол. В таком случае среди поворотов в группе K_3 найдется поворот на наименьший положительный угол. Пусть этот угол равен α°. Докажем, что любой другой поворот, содержащийся в группе, будет кратен α°. Обозначим число градусов в этом повороте через β и найдем такое целое число h, чтобы $h\alpha^\circ \leqslant \beta^\circ < (h+1)\alpha^\circ$, откуда $0 \leqslant \beta^\circ - h\alpha^\circ < \alpha^\circ$. Группа K_3, имея повороты на α° и β°, будет иметь и поворот на $\beta^\circ - h\alpha^\circ$. Но $0 \leqslant \beta^\circ - h\alpha^\circ < \alpha^\circ$, а группа не содержит положительных поворотов меньше чем на α°. Поэтому $\beta^\circ - h\alpha^\circ = 0$, т. е. $\beta^\circ = h\alpha^\circ$. В частности, поскольку группа K_3 содержит поворот на 360°, то для некоторого *целого* n имеем $n\alpha^\circ = 360^\circ$, откуда $\alpha^\circ = \frac{360^\circ}{n}$.

Итак, группа K_3 состоит из поворотов на $0°, \frac{360°}{n}, 2\frac{360°}{n}, \ldots$ $\ldots, (n-1)\frac{360°}{n}$. Придавая n значения 2, 3, 4, ..., мы получим все типы групп K_3. Пример фигур, группа симметрий которых состоит только из поворотов вокруг O на углы, кратные $\frac{360°}{n}$, при $n=19$, $n=3$, приведен на рис. 16.

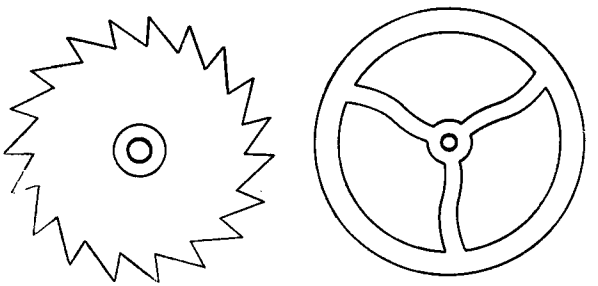


Рис. 16.

4. Группа симметрий K_4 состоит из одних поворотов и содержит сколь угодно малые повороты. Тогда поворот на произвольный угол α можно с любой степенью точности составить из поворотов, принадлежащих группе K_4. Нас здесь интересуют, конечно, только замкнутые фигуры, т. е. такие, которые включают в себя все свои граничные точки (см. главу XVII, § 9). Легко установить, что для замкнутых фигур группа K_4 содержит повороты на любой угол φ. Это — случай направленной круговой симметрии, примером которой может служить окружность, кольцевая полоса и т. п., снабженные некоторым направлением обхода (рис. 17). Здесь при всех допустимых преобразованиях должна совмещаться не только сама фигура, но и направление ее обхода, что исключает отражения относительно прямой.

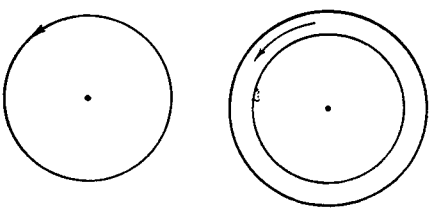


Рис. 17.

Нам остается рассмотреть еще смешанные случаи, когда группа симметрий K содержит и повороты и отражения. Опуская соответствующие доказательства, которые и здесь остаются очень простыми, укажем лишь результат: кроме групп $K_1 - K_4$, существуют еще группы только следующих двух типов.

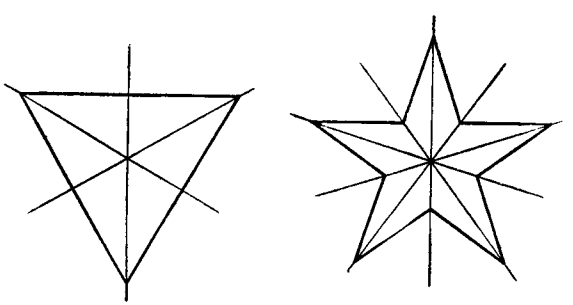


Рис. 18.

5. Группа симметрий K_5 состоит из n отражений относительно прямых, проходящих через O, делящих плоскость на $2n$ равных углов, и

из поворотов на углы, кратные $\frac{360^\circ}{n}$. Такой группой симметрии обладает, например, правильный n-угольник (рис. 18).

6. Группа симметрий K_6 состоит из всех поворотов около O и отражений относительно всевозможных прямых, проходящих через центр O. Это — случай полной круговой симметрии, примером которой может служить симметрия неѝаправленной окружности или ненаправленного кругового кольца.

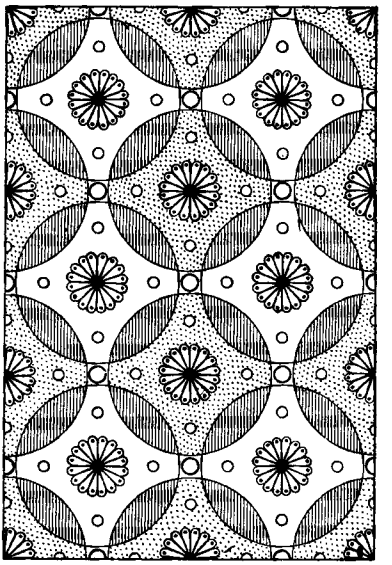


Рис. 19.

Группы симметрий бесконечных плоских фигур. Нахождение всевозможных групп симметрий бесконечных плоских фигур — задача более сложная. Конечно, практически нам никогда не бывает дана вся бесконечная плоскость. Однако кусок плоскости часто бывает покрыт столь мелкими фигурами, что сам кусок представляется по сравнению с ними бесконечно большим. Например, гладко отшлифованная плоскость куска стали оказывается покрытой узорами микроскопических размеров. Правильность этих узоров свидетельствует о внутренней однородности структуры металла.

Другим примером могут служить росписи стен и тканей повторяющимися фигурами. Искусство такой росписи — искусство орнамента — широко развито у большинства народов, начиная с древнейших времен и кончая нашими днями. На рис. 19 дан образец египетской росписи потолка, восходящей к середине второго тысячелетия до нашей эры.

Уже для групп симметрий конечных фигур мы были принуждены отдельно рассматривать случаи 1, 2, 3, 5, когда группа симметрий не содержит поворотов на сколь угодно малый угол, и случаи 4, 6, когда в группе есть такие повороты. При изучении групп симметрий бесконечных фигур, особенно в пространственном случае, это разделение на дискретные группы и группы со сколь угодно малыми преобразованиями приобретает еще большее значение. Поэтому мы сначала более точно проведем разграничение между этими случаями.

Группа движений плоскости называется *дискретной*, если каждую точку плоскости можно окружить таким кругом, что каждое движение из группы либо оставляет данную точку неподвижной, либо выводит ее сразу за пределы взятого круга.

Как и выше, можно найти все дискретные группы движений плоскости. Все эти группы являются группами симметрий плоских фигур. Здесь естественно различаются три типа дискретных групп симметрий:

I. На плоскости существует точка, остающаяся неподвижной при всех преобразованиях симметрии. Этот тип содержит перечисленные выше группы K_1, K_2, K_3 и K_5.

II. На плоскости не существует неподвижной точки, но существует прямая, совмещающаяся с собой при всех преобразованиях группы. Эта прямая называется осью группы. Группы симметрий этого типа имеют орнаменты, вытянутые в виде бесконечной полосы (бордюры). Таких групп существует всего семь:

1. Группа симметрий L_1, состоящая только из переносов на расстояния, кратные некоторому отрезку a.

2. Группа L_2, получающаяся из L_1 присоединением вращения на 180° относительно одной из точек оси группы.

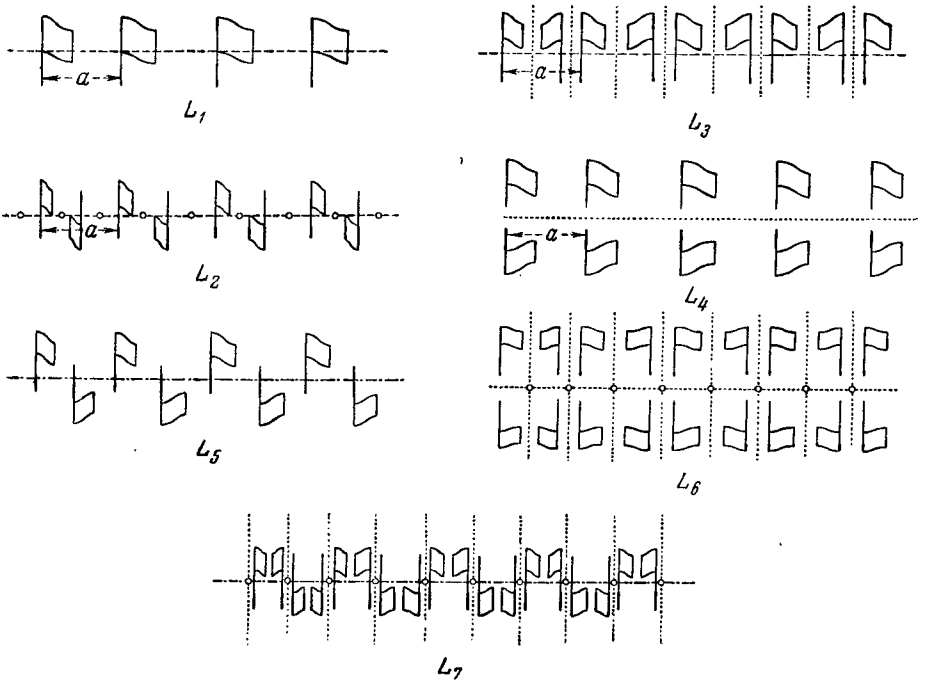
3. Группа L_3, получающаяся из L_1 присоединением отражения относительно прямой, перпендикулярной к оси группы.

4. Группа L_4, получающаяся из L_1 присоединением отражения относительно оси.

5. Группа L_5, получающаяся из L_1 присоединением переноса на отрезок $\frac{a}{2}$ соединенного с отражением относительно оси.

6. Группа L_6, получающаяся из L_4 присоединением отражения относительно некоторой прямой, перпендикулярной оси группы.

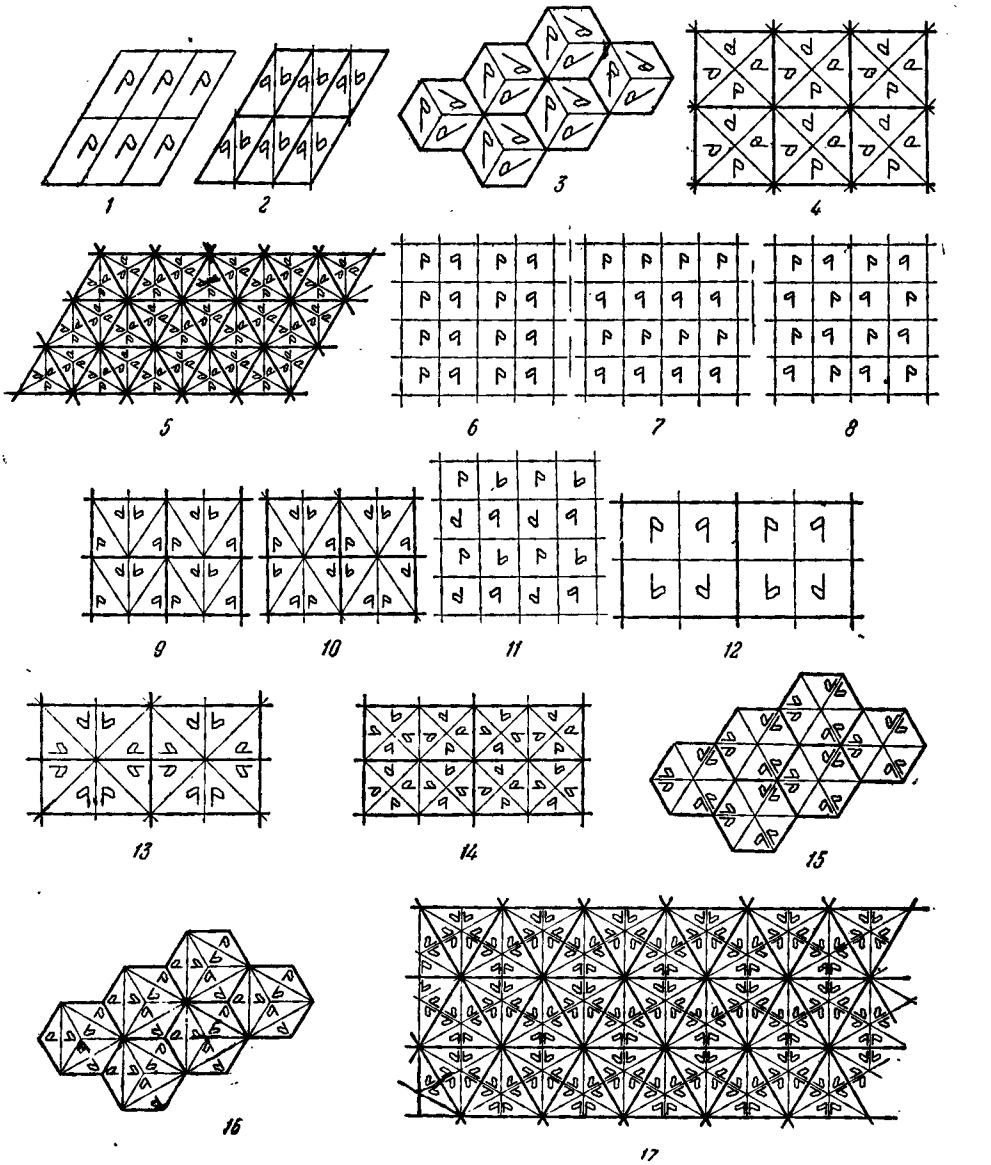
Таблица 2



7. Группа L_7, получающаяся из L_5 присоединением отражения относительно некоторой прямой, перпендикулярной к оси группы.

Табл. 2 дает схемы «бордюров», соответствующих каждой из групп $L_1 - L_7$.

Таблица 3



III. На плоскости не существует ни точки, ни прямой, совмещающихся с собой при всех преобразованиях группы. Группы этого типа называются *плоскими федоровскими группами*. Они являются группами симметрий бесконечных плоских орнаментов. Их существует всего 17:

пять состоит только из движений 1-го рода, двенадцать — из движений 1 и 2-го рода.

В табл. 3 даны схемы орнаментов, соответствующих каждой из 17 плоских федоровских групп; при этом каждая группа состоит из тех и только тех движений, которые любой флаг, начерченный на чертеже, переводят в любой другой флаг того же чертежа.

Интересно отметить, что мастера орнамента практически открыли орнаменты со всеми возможными группами симметрий; на долю теории групп выпало лишь доказать отсутствие других видов.

Кристаллографические группы. В 1890 г. знаменитый русский кристаллограф и геометр Е. С. Федоров, решил теоретико-групповыми методами одну из основных задач кристаллографии: задачу классификации правильных пространственных систем точек. Это было первым случаем непосредственного применения теории групп к решению больших задач естествознания и оказало существенное влияние на развитие теории групп.

Кристаллические тела обладают той особенностью, что составляющие их атомы образуют в пространстве в некотором смысле правильную систему. Рассмотрим те движения пространства, которые переводят точки системы снова в точки системы. Эти движения образуют группу, свойства которой позволяют более точно сформулировать и само понятие правильной системы точек.

Систему точек пространства называют *правильной пространственной системой точек*, если

1) любую точку системы можно перевести в любую другую ее точку посредством движения, совмещающего систему с собой;

2) никакой шар конечного радиуса не содержит бесконечного числа точек системы;

3) существует такое положительное число r, что всякий шар радиуса r содержит по меньшей мере одну точку системы.

Задача изучения строения кристаллических тел оказывается тесно связанной с классификацией правильных пространственных систем точек, которая в свою очередь связана с классификацией дискретных групп движений пространства. Подобно плоскому случаю, группа движений H пространства называется *дискретной*, если около каждой точки A пространства можно описать такой шар положительного радиуса r с центром в A, что каждое движение, входящее в H, или оставляет точку A на месте, или выводит ее за пределы шара.

Можно показать, что совокупность движений пространства, совмещающих с собою данную правильную пространственную систему точек, является обязательно дискретной группой, причем все точки системы можно получить из любой фиксированной точки системы, сдвигая ее при помощи преобразований этой группы. Обратно, если известна некоторая дискретная группа H, то, беря в пространстве произвольную

точку A и сдвигая ее при помощи всевозможных движений, входящих в H, мы получим систему точек, обладающую свойствами 1, 2. Путем несложных дополнительных требований можно из дискретных групп выделить те группы, которые для подходяще выбранных точек A дают настоящие правильные пространственные системы точек, т. е. системы точек со всеми тремя свойствами 1, 2, 3. Такие дискретные группы носят название *федоровских* или *кристаллографических* групп. Из сказанного видно, что нахождение федоровских групп есть первый и важнейший шаг в исследовании правильных пространственных систем точек. Оказалось, что для целей естествознания необходимо рассматривать не только группы, составленные лишь из собственных движений, но также группы, содержащие и собственные и несобственные движения (т. е. включающие отражение). Число федоровских групп, составленных только из собственных движений, значительно меньше числа федоровских групп, составленных из собственных и несобственных движений, и только в последнем, более общем случае многообразие получаемых правильных пространственных систем точек действительно исчерпывает все богатство встречающихся в природе строений кристаллических тел.

Интересно отметить, что лишь теория групп позволила разобраться в этом исключительном богатстве возможностей в отличие от упоминавшегося выше плоского случая.

Сложность пространственной задачи сравнительно с плоской видна из следующей таблицы:

Число пространственных федоровских групп

Групп, содержащих только движения 1-го рода...	65
Групп, содержащих также движения 2-го рода...	165
Всего...	230

Подробный вывод и перечисление всех федоровских групп в пространстве еще и в настоящее время требуют нескольких десятков страниц текста. Поэтому мы ограничимся сообщением только этих количественных результатов, отсылая интересующихся читателей к специальной литературе[1].

Современное развитие кристаллографии сделало необходимым дальнейшее развитие понятия симметрии. Новые возможности и пути этого намечены в книге кристаллографа акад. А. В. Шубникова «Симметрия и антисимметрия конечных фигур», Изд-во АН СССР, 1951.

[1] Подробный вывод плоских дискретных групп движений 1-го рода изложен в книге Д. Гильберта и С. Э. Кон-Фоссена «Наглядная геометрия» (М.—Л., 1951). Вывод пространственных кристаллографических групп можно найти в основном труде Е. С. Федорова «Симметрия правильных систем фигур» (Е. С. Федоров. Основные работы. Изд-во АН СССР, 1949), а также в книге Б. Делоне, Н. Падурова и А. Александрова «Математические основы структурного анализа кристаллов» (М.—Л., 1934) или С. А. Богомолова «Вывод правильных систем по методу Федорова» (изд-во Кубуч, 1934).

§ 5. ГРУППЫ ГАЛУА

Результаты, изложенные на предыдущих страницах, дают некоторое представление о роли, которую сыграла теория групп в решении задачи о классификации кристаллических форм. Однако не эта задача послужила причиной возникновения теории групп. Приблизительно на сто лет раньше Лагранжем была замечена связь между свойствами симметрии корней алгебраического уравнения и возможностью решения уравнения в радикалах. В трудах знаменитых математиков первой трети прошлого века Абеля и Галуа эта связь была глубоко исследована, что привело их к решению знаменитой проблемы об условиях разрешимости алгебраических уравнений в радикалах. Это решение целиком опиралось на тонкие рассмотрения свойств групп подстановок и явилось фактически началом существования теории групп.

Изучение связей между свойствами алгебраических уравнений и свойствами групп составляет ныне предмет обширной теории, известной под именем теории Галуа.

Понятие об истории вопроса и о значении теории Галуа изложено в главе IV (том 1). Поскольку, однако, теория Галуа сыграла решающую роль в развитии теории групп, мы здесь снова приведем основные факты этой теории, но уже в виде, наиболее удобном для освещения самой теории групп. Доказательства этих фактов требуют многочисленных вспомогательных понятий и будут нами опущены.

Группа алгебраического уравнения. Пусть дано уравнение n-й степени

$$x^n + a_1 x^{n-1} + \ldots + a_n = 0, \tag{6}$$

коэффициенты которого считаются данными величинами, например, некоторыми комплексными числами. Совокупность величин, которые можно получить из коэффициентов уравнения при помощи конечного числа действий сложения, вычитания, умножения и деления, называется *основным полем* или областью рациональности уравнения.

Например, если уравнение имеет рациональные коэффициенты, то область рациональности будет состоять из всех рациональных чисел; если же уравнение имеет вид $x^2 + \sqrt{2}\,x + 1 = 0$, то область рациональности состоит из всех чисел вида $a + b\sqrt{2}$, где a, b — рациональные числа.

Обозначим теперь через $\xi_1, \ldots, \xi_n$ корни данного уравнения. Совокупность величин, которые можно получить при помощи конечного числа действий сложения, вычитания, умножения и деления, исходя из корней $\xi_1, \ldots, \xi_n$, называется *полем разложения* уравнения. Так, например, поле разложения уравнения $x^2 + 1 = 0$ есть совокупность комплексных чисел $a + bi$ с рациональными a, b, а поле разложения упомянутого выше уравнения $x^2 + \sqrt{2}\,x + 1 = 0$ есть совокупность чисел вида $a + bi + c\sqrt{2} + di\sqrt{2}$, где a, b, c, d — рациональные числа.

В силу формул Виета коэффициенты уравнения получаются из его корней при помощи операций сложения и умножения, поэтому поле разложения уравнения всегда содержит его основное поле. Иногда эти поля совпадают.

Взаимно однозначное отображение A поля разложения на себя называется автоморфизмом поля разложения относительно основного поля, если для каждой пары элементов поля разложения их сумма переходит в сумму, произведение — в произведение, а каждый элемент основного поля переходит сам в себя. Указанные свойства можно записать формулами

$$(a+b)A=aA+bA,\quad (ab)A=aA\cdot bA,\quad \alpha A=\alpha \qquad (a,\ b\in K,\quad \alpha\in P), \tag{7}$$

где aA — тот элемент, в который переходит элемент a при отображении A; P — основное поле; K — поле разложения.

Согласно упоминавшемуся на стр. 255 общему принципу совокупность всех автоморфизмов поля разложения относительно основного поля является группой. Эта группа и называется *группой Галуа* данного уравнения.

Чтобы составить себе несколько более конкретное представление о группе Галуа, заметим прежде всего, что автоморфизмы из группы Галуа переводят корни данного уравнения снова в корни этого же уравнения. Действительно, если x — корень уравнения (6), то, действуя на обе части этого уравнения автоморфизмом A и пользуясь свойствами (7), получим

$$(xA)^n+a_1A\,(xA)^{n-1}+\ldots+a_nA=0\cdot A;$$

так как $0\cdot A=0$, $a_iA=a_i$, то отсюда имеем

$$(xA)^n+a_1(xA)^{n-1}+\ldots+a_n=0,$$

что и требовалось. Следовательно, каждый автоморфизм A вызывает определенную подстановку множества корней уравнения. С другой стороны, зная эту подстановку, мы знаем и сам автоморфизм, поскольку все элементы поля разложения получаются из корней только при помощи арифметических операций. Это показывает, что вместо группы автоморфизмов можно рассматривать соответствующую ей группу подстановок корней уравнения. Отсюда следует, в частности, что все группы Галуа конечные.

Найти группу Галуа какого-либо уравнения — задача обычно сложная и лишь в отдельных случаях группа Галуа находится сравнительно просто. Рассмотрим, например, уравнение (6) с буквенными коэффициентами $a_1,\ldots,a_n$. Основное поле этого уравнения составляют рациональные дроби от коэффициентов, т. е. дроби, числители и знаменатели которых являются многочленами от $a_1,\ldots,a_n$. Поле разложения составляют рациональные дроби от корней уравнения $\xi_1,\ldots,\xi_n$,

связанных с коэффициентами формулами

$$\begin{aligned} -a_1 &= \xi_1 + \xi_2 + \ldots + \xi_n, \\ a_2 &= \xi_1\xi_2 + \xi_1\xi_3 + \ldots + \xi_{n-1}\xi_n, \\ &\cdots\cdots\cdots\cdots\cdots \\ (-1)^n a_n &= \xi_1\xi_2 \ldots \xi_n. \end{aligned} \tag{8}$$

Поскольку уравнение (6) «общее», мы можем считать его корни независимыми переменными. Тогда всякая подстановка этих корней будет вызывать автоморфизм поля разложения. Формулы (8) показывают, что при любом таком автоморфизме коэффициенты переходят сами в себя, а вместе с ними сами в себя переходят и все рациональные дроби от них. Таким бразом, группа Галуа общего уравнения n-й степени по существу является симметрической группой всех подстановок n букв.

Можно указать также уравнения с численными коэффициентами, имеющие симметрическую группу своей группой Галуа. Доказано, например, что группа Галуа уравнения

$$1 - \frac{n}{1}x + \frac{n(n-1)}{1\cdot 2}\cdot\frac{1}{1\cdot 2}x^2 - \frac{n(n-1)(n-2)}{1\cdot 2\cdot 3}\cdot\frac{1}{1\cdot 2\cdot 3}x^3 + \ldots + (-1)^n\frac{1}{1\cdot 2\ldots n}x^n = 0 \tag{9}$$

при любом n есть симметрическая группа подстановок степени n.

Вообще известны способы построения уравнений с любой наперед заданной группой в качестве группы Галуа, но при условии, что коэффициенты можно брать произвольными. Если же требовать построения уравнений, имеющих непременно рациональные коэффициенты, то такое построение в настоящее время известно лишь для отдельных типов групп. Значительного успеха в этом направлении добился советский математик И. Р. Шафаревич, нашедший способы построения уравнений с рациональными коэффициентами, имеющими своей группой Галуа любую наперед заданную разрешимую группу. В общем же случае эта задача остается пока нерешенной.

Разрешимость уравнений в радикалах. Группа Галуа уравнения характеризует, как это видно из определения, внутреннюю симметрию корней уравнения. Оказалось, что все наиболее существенные вопросы, касающиеся возможности свести решение заданного уравнения к решению уравнений низших степеней, а также многие другие, могут быть сформулированы в виде вопросов о строении группы Галуа, а группа Галуа каждого уравнения n-й степени есть некоторая группа подстановок n-й степени, т. е. объект вполне конечный, все соотношения в котором, по меньшей мере теоретически, можно найти хотя бы путем проб.

Изучение группы Галуа представляет собой замечательный метод решения проблем, относящихся к алгебраическим уравнениям высших степеней. Например, можно доказать, что уравнение разрешимо в радикалах в том и только в том случае, если разрешима его группа Галуа (определение разрешимой группы см. § 3, стр. 268). Уже упоминалось,

что симметрические группы 2, 3 и 4-й степени разрешимы. Это находится в полном согласии с общеизвестным фактом разрешимости в радикалах уравнений 2, 3 и 4-й степени. Группы Галуа «общих» уравнений 5, 6-й и т. д. степеней суть симметрические группы тех же степеней. Но эти группы неразрешимы. Отсюда следует, что общие уравнения выше 4-й степени не могут быть разрешены в радикалах.

К числу уравнений, не разрешимых в радикалах, относятся и уравнения (9) при $n>4$, поскольку их группа Галуа также симметрическая.

§ 6. ОСНОВНЫЕ ПОНЯТИЯ ОБЩЕЙ ТЕОРИИ ГРУПП

В XIX в. теория групп развивалась преимущественно как теория групп преобразований. Однако с течением времени становилось все более ясным, что наиболее существенные из полученных результатов зависят лишь от того, что преобразования можно перемножать и что это действие обладает рядом характерных свойств. С другой стороны, были найдены объекты, вовсе не являющиеся преобразованиями, но над которыми можно производить некоторое действие (назовем его по-прежнему умножением), обладающее теми же свойствами, что и в группах преобразований, и к которым главные теоремы теории групп преобразований оказались также применимыми. В силу этого к концу прошлого столетия понятие группы стали применять не только к системам преобразований, но и к системам произвольных элементов.

Общее определение группы. В настоящее время общепринято следующее определение группы: пусть каждой паре рассматриваемых в определенном порядке элементов a, b произвольного множества G сопоставлен вполне определенный элемент с того же множества. Тогда говорят, что на множестве G задана операция или действие. Обычно для операций вводят особые названия: сложение, умножение, композиция. Элемент множества G, отвечающий паре a, b, называется в таком случае соответственно суммой, произведением, результатом композиции элементов a, b и обозначается соответственно $a+b$, ab, $a*b$. Название «сложение» или «умножение» употребляется и в тех случаях, когда рассматриваемая операция не имеет никакого отношения к обычным действиям сложения и умножения чисел.

Множество G вместе с определенной на нем операцией $*$ называется *группой* относительно этой операции, если выполняются следующие групповые аксиомы:

1. Для любых трех элементов x, y, z из G

$$x*(y*z)=(x*y)*z \text{ (закон ассоциативности).}$$

2. Среди элементов G существует элемент e, для которого при любом x из G

$$x*e=e*x=x.$$

3. Для каждого элемента a из G в G существует такой элемент a^{-1}, что

$$a*a^{-1}=a^{-1}*a=e.$$

Элемент e, указанный в аксиоме 2, называется нейтральным элементом группы, а элемент a^{-1}, существование которого утверждает аксиома 3, называется обратным для a. Если групповая операция называется сложением или умножением, то нейтральный элемент называется соответственно нулевым или единичным, а групповые аксиомы принимают вид

1) $x+(y+z)=(x+y)+z$,	1) $x(yz)=(xy)z$,
2) $x+0=0+x=x$,	2) $xe=ex=x$,
3) $x+(-x)=(-x)+x=0$,	3) $xx^{-1}=x^{-1}x=e$.

В предшествующих параграфах рассмотрено много примеров групп. Элементами этих групп были преобразования, а групповым действием служило умножение преобразований. Совокупность чисел 0, ± 1, ± 2, ... относительно операции сложения является также группой, так как сумма целых чисел есть снова целое число, сложение целых чисел ассоциативно; нейтральным элементом является целое число 0, и для каждого числа a среди рассматриваемых чисел есть противоположное число $-a$. Другим примером группы может служить совокупность всех действительных чисел (за исключением нуля) относительно умножения. В самом деле, произведение любых двух, отличных от нуля, действительных чисел есть действительное число, отличное от нуля; действие умножения действительных чисел ассоциативно; нейтральный элемент существует и равен 1; у каждого ненулевого действительного числа a есть обратное число $a^{-1}=\frac{1}{a}$. Число аналогичных примеров можно неограниченно увеличить.

Хотя групповая операция может называться различно, мы условимся в дальнейшем почти всегда называть ее умножением. Понятия подгруппы, степеней элемента группы, циклической группы, порядка элемента группы определяются так же, как и для групп преобразований, и мы их повторять не будем (см. § 3). Отметим лишь, что элемент a группы G называется сопряженным с элементом b, если в G найдется такой элемент x, что $b=x^{-1}ax$. Так как $a=a^{-1}aa$, то каждый элемент группы сопряжен с самим собой. Далее, из $b=x^{-1}ax$, очевидно, следует $xbx^{-1}=a$, или $a=(x^{-1})^{-1}bx^{-1}$, т. е. если элемент a сопряжен с b, то b сопряжен с a. Наконец, если $b=x^{-1}ax$ и $c=y^{-1}by$, то

$$c=y^{-1}x^{-1}axy=(xy)^{-1}a(xy).$$

Следовательно, два элемента, сопряженные с третьим, сопряжены между собой. Эти свойства показывают, что в группе все элементы распадаются на различные классы сопряженных друг с другом элементов. Впрочем, если группа коммутативна, т. е. $xy=yx$ для любых x и y, то сопряженные элементы совпадают и каждый класс сопряженных элементов оказывается состоящим только из одного элемента.

Изоморфизм. В понятии группы можно различать две стороны. Чтобы задать группу, нужно: 1) указать, какие объекты являются ее элементами, и 2) указать закон перемножения элементов. Сообразно этому и изучение свойств групп можно производить с разных точек зрения. Можно изучать связи между индивидуальными свойствами элементов группы и их совокупностей и свойствами их по отношению к групповой операции. Такой точки зрения часто держатся при изучении отдельных конкретных групп, например при изучении свойств группы движений пространства или плоскости. Однако можно изучать и те свойства групп, которые целиком выражаются через свойства групповой операции. Эта точка зрения характерна для *абстрактной* или *общей* теории групп. Более отчетливо она может быть выражена при помощи понятия изоморфизма.

Две группы называются *изоморфными*, если элементы одной из них можно так сопоставить с элементами другой, что произведению произвольных элементов первой группы будет отвечать произведение соответствующих элементов второй группы. Взаимно однозначное соответствие между элементами двух групп, обладающее указанными свойствами, называется *изоморфизмом*.

Легко видеть, что элементы двух групп, отвечающие друг другу при изоморфном соответствии, будут обладать одинаковыми свойствами по отношению к групповой операции. Так, при изоморфном соответствии нейтральный элемент, взаимно обратные элементы, элементы данного порядка n, подгруппы одной группы переходят соответственно в нейтральный элемент, взаимно обратные элементы, элементы того же порядка n, подгруппы другой группы. Поэтому можно сказать, что абстрактная теория групп изучает лишь те свойства групп, которые сохраняются при изоморфных отображениях. Например, с точки зрения абстрактной теории групп группа всех подстановок четырех элементов и группа собственных и несобственных движений пространства, переводящих в себя фиксированный правильный четырехгранник, обладают одинаковыми свойствами, так как они изоморфны. Действительно, рассматриваемые движения переводят вершины четырехгранника снова в его вершины. Число этих движений равно 24. Сопоставляя с каждым движением вызываемую им перестановку вершин, мы получим взаимно однозначное соответствие между элементами обеих групп, которое и будет искомым изоморфизмом.

Замечательный пример изоморфного соответствия дает теория логарифмов. Ставя каждому положительному действительному числу в соответствие логарифм этого числа, мы получим взаимно однозначное отображение множества действительных положительных чисел на множество всех положительных и отрицательных действительных чисел. Соотношение $\log(xy) = \log x + \log y$ показывает, что установленное отображение является изоморфным отображением группы положитель-

ных действительных чисел относительно умножения на группу всех действительных чисел относительно сложения. Практическая важность этого изоморфизма общеизвестна.

Примерами неизоморфных групп могут служить конечные группы различных порядков.

Как уже было сказано выше, абстрактная группа определяется законом умножения элементов, независимо от их природы, так что различные изоморфные между собой конкретно заданные группы можно рассматривать как модели одной и той же абстрактной группы.

Абстрактную группу можно задавать различными способами, из которых наиболее естественным, по крайней мере для конечных групп, является задание посредством «таблицы умножения».

Для группы порядка n, элементы которой записаны в каком-либо порядке, такая таблица умножения состоит из квадрата, разделенного на n строк и n столбцов. В клетке, лежащей в i-й строке и j-м столбце, обозначается элемент, являющийся произведением элемента с номером i на элемент с номером j. Такая таблица умножения для конечной группы иногда называется ее квадратом Кэли.

Однако практически задание группы посредством таблицы умножения почти не употребляется ввиду большой громоздкости.

Существуют и другие способы задания абстрактной группы. С одним из них — заданием группы при помощи производящих элементов и определяющих соотношений — мы еще познакомимся. Однако чаще всего абстрактную группу определяют заданием изоморфной ей конкретной группы, в частности группы преобразований.

Возникает естественный вопрос, можно ли любую абстрактную группу рассматривать как группу преобразований. Ответ дает следующая теорема: каждая группа G изоморфна некоторой группе преобразований множества ее элементов.

Действительно, пусть g — фиксированный элемент из G. Обозначим через A_g то преобразование множества элементов G, при котором каждому элементу x из G отвечает элемент xg. Преобразование A_g взаимно однозначное, так как уравнение

$$xA_g = xg = a$$

при любом данном a имеет единственное решение $x = ag^{-1}$. С другой стороны, произведению элементов группы gh отвечает произведение соответствующих преобразований $A_g A_h$, ибо

$$xA_{gh} = x(gh) = (xg)h = (xA_g)A_h = x(A_g A_h).$$

Нейтральному элементу e группы G отвечает единичное, а обратному элементу g^{-1} обратное преобразование. Поэтому совокупность Γ всех преобразований, отвечающих элементам группы G, является группой преобразований, изоморфной G. Легко убедиться, что если число элементов G больше 2, то совокупность Γ не исчерпывает всех преобра-

зований множества G и является лишь подгруппой «симметрической» группы всех преобразований этого множества.

Инвариантные подгруппы и фактор-группы. Пусть P и Q — произвольные совокупности элементов какой-либо группы G. Произведением совокупности P на совокупность Q, символически PQ, называется множество тех элементов группы G, которые могут быть представлены в виде произведения некоторого элемента из P на некоторый элемент из Q. В частности, произведение gP, где g — элемент группы G, есть совокупность произведений элемента g на каждый элемент множества P.

Подгруппа H группы G называется *инвариантной подгруппой* или *нормальным делителем* группы G, если $gH = Hg$ для любого g из G. Совокупности вида gH и Hg, где H — произвольная подгруппа, называются соответственно *правым и левым смежными классами* группы G по подгруппе H, содержащими элемент g. Таким образом, можно сказать, что инвариантные подгруппы вполне характеризуются тем свойством, что для них левый и правый смежные классы, отвечающие одному и тому же элементу, совпадают.

Если H — инвариантная подгруппа, то произведение двух смежных классов будет, как легко доказать, снова смежным классом, именно: $aH \cdot bH = ab \cdot H$. Подгруппа H сама по себе является смежным классом, отвечающим единице или любому своему элементу h, так как $hH = H$. Умножение смежных классов ассоциативно

$$(aH \cdot bH)\, cH = (ab \cdot c)\, H = (a \cdot bc)\, H = aH\, (bH \cdot cH).$$

Подгруппа H играет роль нейтрального элемента в этом умножении: $H \cdot aH = eH \cdot aH = (ea)\, H = aH$, аналогично $aH \cdot H = aH$. Смежный класс $a^{-1}H$ является обратным относительно aH, так как $aH \cdot a^{-1}H = aa^{-1}H = H$. Следовательно, рассматривая каждый смежный класс по инвариантной подгруппе в качестве элемента нового множества, мы видим, что это множество будет группой относительно операции умножения смежных классов. Эта группа называется *фактор-группой* группы G по инвариантной подгруппе H и обозначается G/H.

Легко установить, что для конечных групп каждый смежный класс по любой подгруппе H содержит столько различных элементов, сколько их содержит сама подгруппа H, и что разные смежные классы общих элементов не имеют. Отсюда следует, что число смежных классов конечной группы G по какой-либо ее подгруппе H равно порядку G, деленному на порядок H, откуда вытекает важная теорема Лагранжа о том, что порядок любой подгруппы конечной группы является делителем порядка группы.

Из определения инвариантной подгруппы видно, что в абелевых группах всякая подгруппа является инвариантной. Другой крайний случай представляют так называемые *простые группы*, ни одна подгруппа которых, отличная от единичной и от самой группы, не

является инвариантной. Помимо абелевых и простых групп, важное значение имеют также разрешимые группы, определение которых было уже дано в § 3. Можно показать, что разрешимые группы обладают конечной цепочкой инвариантных подгрупп G, G_1, G_2, ..., G_k, первая из которых совпадает с заданной группой G; каждая следующая содержится в предыдущей; последняя подгруппа является единичной, а все фактор-группы G/G_1, G_1/G_2, ..., G_{k-1}/G_k являются абелевыми.

Гомоморфизм. Понятие фактор-группы весьма тесно связано с основным для всей теории групп понятием гомоморфного отображения.

Однозначное отображение совокупности элементов группы G на совокупность элементов группы H называется *гомоморфизмом* или *гомоморфным отображением*, если произведению каждых двух элементов первой группы отвечает произведение соответствующих элементов второй.

Таким образом, обозначая для каждого элемента x из группы G через x' соответствующий элемент группы H, гомоморфное отображение можно характеризовать свойством

$$(x_1x_2)' = x_1'x_2'.$$

Из определений гомоморфизма и изоморфизма видно, что изоморфное отображение обязательно взаимно однозначное, в то время как гомоморфное отображение однозначно только в одну сторону: каждому элементу группы G отвечает единственный элемент группы H, но различные элементы G могут иметь один и тот же образ в H. В известном смысле можно утверждать, что при изоморфном отображении группа H является точной копией группы G, а при гомоморфном отображении, при переходе от G к H, различия между некоторыми элементами G стираются, некоторые элементы как бы «склеиваются» в один элемент H. Однако эта «грубость» гомоморфного отображения не есть недостаток, а, наоборот, является большим преимуществом, позволяющим употреблять гомоморфное отображение в качестве мощного средства для исследования свойств групп.

Во многих положениях, связанных с преобразованиями, гомоморфное отображение появляется само собой. Рассмотрим, например, группу симметрий правильного тетраэдра (рис. 20). Эта группа изоморфна симметрической группе подстановок четырех элементов, ибо существует одно и только одно движение (1-го или 2-го рода), переводящее вершины A_1, A_2, A_3, A_4 в любое другое заданное расположение.

Рассмотрим теперь прямые l_1, l_2, l_3, соединяющие середины противоположных ребер. Каждое движение, совмещающее тетраэдр с собой, порождает некоторую подстановку l_1, l_2, l_3 и всякая подстановка l_1, l_2, l_3 порождается некоторой симметрией тетраэдра. Ясно, что произведению преобразований тетраэдра соответствует произведение подстановок прямых l_1, l_2, l_3. По рис. 20 легко проследить, как естественным образом осуществляется при этом гомоморфное отображение

симметрической группы подстановок четырех элементов A_1, A_2, A_3, A_4 на симметрическую группу подстановок трех элементов l_1, l_2, l_3. Без труда можно найти элементы «большей» группы, которые «склеиваются» при этом гомоморфизме.

Рассмотрим еще несколько примеров. Совокупность всех подстановок n элементов является при $n>2$ некоммутативной группой. С другой стороны, числа $+1$ и -1 относительно умножения также образуют группу. Поставим теперь в соответствие каждой четной подстановке каких-либо n элементов число $+1$, а каждой нечетной подстановке число -1. Это дает гомоморфное отображение симметрической группы подстановок n элементов на группу $\{+1, -1\}$, так как, согласно § 3, произведение подстановок одинаковой четности есть подстановка четная, а произведение подстановок различной четности есть подстановка нечетная.

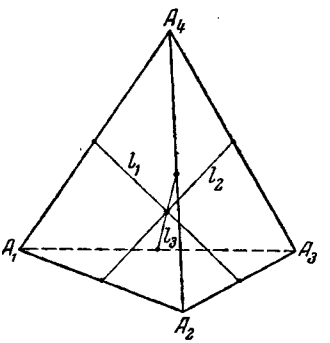


Рис. 20.

Другой пример: если каждому действительному числу $x\neq 0$ поставить в соответствие его абсолютную величину $|x|$, то возникающее отображение группы положительных и отрицательных действительных чисел относительно умножения (нуль исключается) на группу только положительных действительных чисел относительно умножения будет гомоморфизмом, поскольку $|xy|=|x||y|$.

Выше было отмечено, что на плоскости каждое движение 1-го рода A может быть представлено в виде произведения подходящего поворота V_A вокруг фиксированной точки O и некоторого параллельного переноса D_A. Повороты вокруг точки O образуют группу. Поэтому соответствие $A\to V_A$ однозначно отображает группу плоских движений 1-го рода на группу поворотов плоскости вокруг точки O. Покажем, что это отображение является гомоморфизмом. Из разложений $A=V_AD_A$, $B=V_BD_B$ следует

$$AB=V_AD_AV_BD_B=(V_AV_B)(V_B^{-1}D_AV_BD_B).$$

Первая скобка есть поворот вокруг O, а вторая — произведение преобразованного переноса $V_B^{-1}D_AV_B$ на перенос D_B и, следовательно, является переносом. Это показывает, что произведению движений AB отвечает произведение соответствующих поворотов V_AV_B, т. е. что рассматриваемое отображение есть гомоморфизм.

Наконец, докажем, что фактор-группа G/N произвольной группы G по любому ее нормальному делителю N есть гомоморфный образ группы G.

Действительно, поставив каждому элементу g группы G в соответствие смежный класс gN, содержащий g, мы получаем искомое

гомоморфное отображение G на G/N, так как произведению gh отвечает смежный класс ghN, равный произведению смежных классов gN и hN, отвечающих элементам g и h.

Возвращаясь к общим свойствам гомоморфных отображений, покажем, что нейтральный элемент при любом гомоморфизме переходит в нейтральный элемент и что взаимно обратные элементы переходят во взаимно обратные же.

В самом деле, если e — нейтральный элемент группы G, e' — его образ в H, то из $ee=e$ следует $e'e'=e'$, откуда, обозначая через ε нейтральный элемент группы H, получим: $e'=e'e'^{-1}=\varepsilon$. Первое утверждение доказано. Пусть теперь x и y — взаимно обратные элементы в G, а x' и y' — их образы в H. Из $xy=e$ следует $x'y'=e'=\varepsilon$, т. е. x и y — взаимно обратные элементы в H, и, значит,

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

Доказанные утверждения позволяют легко найти образ любого произведения элементов из G. Например,

$$(ab^{-1}c^{-1}dh^{-1})'=a'(b^{-1'})(c^{-1})'d'(h^{-1})'=a'b'^{-1}c'^{-1}d'h'^{-1}.$$

Следующая теорема лежит в основе всей теории гомоморфных отображений.

При гомоморфном отображении произвольной группы G на группу H совокупность N элементов группы G, отображающихся в нейтральный элемент e' группы H, является инвариантной подгруппой в G; совокупность элементов G, отображающихся в произвольный фиксированный элемент группы H, является смежным классом G по N, а устанавливающееся таким образом взаимно однозначное соответствие между смежными классами G по N и элементами группы H есть изоморфизм между H и фактор-группой G/N.

Докажем эту теорему. Пусть a, b — произвольные элементы из N. Это означает, что $a'=b'=e'$, где штрихом, как и ранее, обозначены образы элементов G в H. Но тогда

$$(ab)'=a'b'=e'e'=e',$$
$$(a^{-1})'=a'^{-1}=e'^{-1}=e',$$

т. е. ab и обратные элементы a^{-1}, b^{-1} принадлежат N, и, следовательно, совокупность N является группой. Далее, для произвольного элемента g из G

$$(g^{-1}ag)'=g'^{-1}a'g'=g'^{-1}e'g'=g'^{-1}g'=e',$$

т. е. $g^{-1}ag$ входит в N при любом g из G и любом a из N, а из этого, очевидно, следует, что N — инвариантная подгруппа. Первое утверждение теоремы доказано.

Для доказательства второго утверждения возьмем в группе G произвольный элемент g и рассмотрим совокупность U всех тех элементов u из G, образ которых u' совпадает с образом g' элемента g. Пусть $u \in gN$, т. е. $u = gn$, где $n \in N$, тогда $u' = g'n' = g'e' = g'$. Следовательно, $gN \subset U$. Обратно, если $u' = g'$, то $(g^{-1}u)' = g'^{-1}u' = g'^{-1}g' = e'$, т. е. $g^{-1}u = n$, где n — элемент из N. Отсюда $u = gn$ и, значит, $U \subset gN$. Из $gN \subset U$, $U \subset gN$ следует: $U = gN$.

Наконец, третье утверждение теоремы очевидно: произвольным смежным классам gN, hN из фактор-группы G/N отвечают в H элементы g', h', а произведению классов, в силу формулы

$$gN \cdot hN = ghN,$$

отвечает $(gh)' = g'h'$, что и требовалось доказать.

Теорема о гомоморфизмах показывает, что каждый гомоморфный образ H группы G изоморфен соответствующей фактор-группе G/H. Таким образом, с точностью до изоморфизма все гомоморфные образы заданной группы G исчерпываются ее различными фактор-группами.

§ 7. НЕПРЕРЫВНЫЕ ГРУППЫ

Группы Ли. Непрерывные группы преобразований. Успех, выпавший на долю теории групп в решении алгебраических уравнений высших степеней, побудил математиков середины прошлого века попытаться применить теорию групп к решению уравнений других видов, в первую очередь к решению дифференциальных уравнений, играющих столь большую роль в приложениях математики. Эта попытка увенчалась успехом. Хотя группы в дифференциальных уравнениях заняли совершенно иное место, нежели в теории алгебраических уравнений, исследования по применениям теории групп к решению дифференциальных уравнений привели к существеннейшему расширению самого понятия группы и созданию новой теории так называемых непрерывных групп и групп Ли, оказавшихся чрезвычайно важными для развития самых разнообразных отделов математики.

В то время как группы алгебраических уравнений состоят лишь из конечного числа преобразований, построенные аналогичным образом группы дифференциальных уравнений оказались бесконечными. Кроме того, оказалось, что преобразования, принадлежащие к группе дифференциального уравнения, возможно задать посредством конечной системы параметров, меняя численные значения которых, можно получить все преобразования группы. Пусть, например, все преобразования группы определяются значениями параметров $a_1, a_2, \ldots, a_r$. Придавая этим параметрам значения $x_1, x_2, \ldots, x_r$, мы получим некоторое преобразование X; придавая тем же параметрам новые значения $y_1, y_2, \ldots, y_r$, получим другое преобразование Y. По условию произведе-

ние этих преобразований $Z=XY$ входит в группу и, значит, получается при определенных новых значениях параметров $z_1, z_2, \ldots, z_r$. Значения z_i будут зависеть от $x_1, x_2, \ldots, x_r, y_1, y_2, \ldots, y_r$, т. е. будут некоторыми функциями от них

$$\begin{aligned}
z_1 &= \varphi_1(x_1, x_2, \ldots, x_r; y_1, y_2, \ldots, y_r),\\
z_2 &= \varphi_2(x_1, x_2, \ldots, x_r; y_1, y_2, \ldots, y_r),\\
&\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\\
z_r &= \varphi_r(x_1, x_2, \ldots, x_r; y_1, y_2, \ldots, y_r).
\end{aligned}$$

Группы, элементы которых непрерывно зависят от значений конечной системы параметров, а закон умножения выражается при помощи дважды дифференцируемых функций $\varphi_1, \ldots, \varphi_r$, называются группами Ли, по имени норвежского математика Софуса Ли, впервые исследовавшего эти группы.

В первой половине XIX в. в трудах Н. И. Лобачевского была изложена новая геометрическая система, носящая ныне его имя. Примерно в то же время в самостоятельную геометрическую систему выделилась проективная геометрия; несколько позже была создана геометрия Римана. В результате, ко второй половине XIX в. можно было насчитать уже ряд самостоятельных геометрических систем, с различных точек зрения изучавших «пространственные формы действительного мира» (Энгельс). Охватить все эти геометрические системы с единой точки зрения, сохранив при этом важнейшие качественные отличия их, оказалось возможным при помощи теории групп.

Рассмотрим взаимно однозначные преобразования совокупности точек какого-либо геометрического пространства, не изменяющие тех основных отношений между фигурами, которые изучаются в данной геометрии. Совокупность этих преобразований составляет группу, называемую обычно группой движений или автоморфизмов данной геометрии. Группа движений вполне характеризует данную геометрию, так как если группа движений известна, то соответствующая геометрия может рассматриваться как наука, изучающая те свойства совокупностей точек, которые остаются неизменными при преобразованиях данной группы. Метод классификации различных геометрических систем по их группам движений был выдвинут во второй половине прошлого века в работах Ф. Клейна. Об этом методе и различных геометрических системах было рассказано в главе об абстрактных пространствах. Здесь же мы только отметим, что группы движений всех фактически изучавшихся в прошлом веке геометрических систем оказались группами Ли. В силу этого задача изучения групп Ли приобрела особую важность.

Благодаря обилию связей с самыми различными областями математики и механики, теория групп Ли энергично развивалась от своего основания вплоть до наших дней. При этом оказалось, что некоторые вопросы, не решенные для конечных групп и в настоящее время, были сравнительно быстро разрешены для групп Ли. Так, задача классификации простых конечных групп (т. е. конечных групп, не имеющих нетривиальных инвариантных подгрупп) остается до сих пор мало продвинутой, а соответствующая классификация простых групп Ли была получена Киллингом и Картаном еще в конце прошлого века. Развивая теорию групп Ли, советские математики В. В. Морозов, А. И. Мальцев, Е. Б. Дынкин нашли полное решение важной проблемы классификации простых *подгрупп* групп Ли, долгое время ожидавшей своего решения. В ином направлении развивалась теория групп Ли советскими математиками И. М. Гельфандом и М. А. Наймарком, нашедшими так называемые неприводимые представления простых групп Ли унитарными преобразованиями гильбертова пространства; последняя задача представляет особый интерес для анализа и физики.

Изучение групп Ли осуществляется посредством своеобразного аппарата так называемых «инфинитезимальных групп», или алгебр Ли. Более подробно они будут рассмотрены в § 13.

Топологические группы. Наряду с широким развитием классической теории групп Ли в СССР исключительных успехов достигла более общая теория топологических или непрерывных групп. В отличие от понятия группы Ли, где требуется, чтобы элементы группы определялись конечной системой параметров и чтобы закон умножения выражался при помощи дифференцируемых функций, понятие топологической группы проще и шире. Именно, группа называется *топологической*, если для ее элементов, кроме обычной групповой операции, определено понятие близости, и при этом из близости элементов группы следует близость их произведений и близость обратных элементов.

Первоначально понятие топологической группы оказалось необходимым, чтобы привести в надлежащий порядок многие основные понятия теории групп Ли. Однако впоследствии обнаружилась чрезвычайно большая важность этого понятия и для других отделов математики. Первые работы по теории общих топологических групп относятся к началу 20-х годов нашего века, но фундаментальные результаты, позволившие говорить о возникновении новой дисциплины, были найдены лишь в конце 20-х—начале 30-х годов. Значительная часть их была получена советским математиком Л. С. Понтрягиным, который заслуженно считается одним из создателей современной теории непрерывных групп. Его книга «Непрерывные группы», содержащая первое в мировой литературе сводное изложение теории непрерывных групп, остается основным руководством в этой области уже почти 20 лет.

§ 8. ФУНДАМЕНТАЛЬНЫЕ ГРУППЫ

Во всех рассмотренных в предыдущих параграфах конкретных случаях группы обычно появлялись как группы преобразований тех или иных множеств. Исключение составляли лишь группы чисел относительно сложения и умножения. Теперь мы хотим разобрать важный пример, когда с самого начала группа возникает не как группа преобразований, а именно как некоторая алгебраическая система с одним действием.

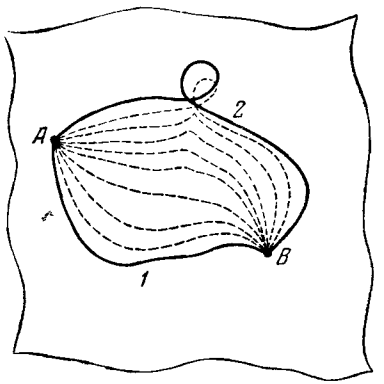


Рис. 21.

Фундаментальная группа. Рассмотрим некоторую поверхность S и на ней подвижную точку M. Заставляя M пробегать на поверхности непрерывную кривую, соединяющую точку A с точкой B, мы получим определенный путь из A в B. Этот путь может пересекать сам себя любое число раз и даже может идти сам по себе на отдельных участках. Чтобы указать путь, мало задать только кривую, по которой перемещается точка M. Нужно еще указать участки, которые эта точка проходит несколько раз, и направление их прохождения. Например, точка может пробегать одну и ту же окружность различное число раз и в различных направлениях, причем все эти круговые пути считаются различными. Два пути с одинаковыми начальными и конечными точками называются эквивалентными, если один из них можно перевести в другой непрерывным изменением. На плоскости или сфере любые два

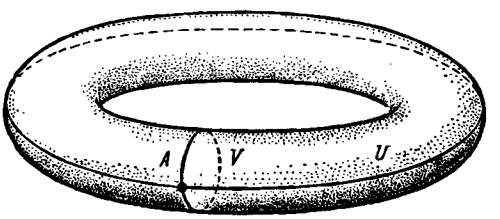


Рис. 22.

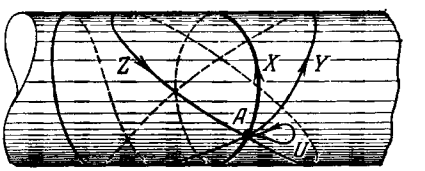


Рис. 23.

пути, соединяющие точку A с точкой B, эквивалентны (рис. 21). Однако на поверхности тора, например, замкнутые пути U и V (рис. 22), выходящие и оканчивающиеся в точке A, не эквивалентны друг другу.

Если вместо тора рассмотреть бесконечно простирающийся в обе стороны круговой цилиндр и на нем взять путь X (рис. 23), то легко сообразить, что любой замкнутый путь с начальной точкой A, проведенный на цилиндре, будет эквивалентен пути вида X^n ($n = 0, \pm 1,$

$\pm 2, \ldots$), где под $X^n (n > 0)$ следует понимать путь X, повторенный n раз; под X^0 — нулевой путь, состоящий лишь из одной точки A, а под X^{-n} — путь X^n, пробегаемый в обратную сторону, например: $Z \sim X^{-1}$, $Y \sim X^2$, $U \sim X^0$ (рис. 23). Этот пример показывает значение понятия эквивалентности путей: в то время как различных замкнутых путей на цилиндре существует необозримое множество, с точностью до эквивалентности все эти пути сводятся только к окружности X, пробегаемой в том или ином направлении достаточное число раз. При $m \neq n$ пути X^m и X^n не эквивалентны.

Возвращаясь к рассмотрению произвольной поверхности, предположим, что нам заданы на ней два пути — путь U, ведущий из точки A в точку B, и путь V, ведущий из B в точку C. Тогда, заставляя сначала точку пробежать путь AB, а затем BC, мы получим путь AC, который естественно назвать произведением путей $U = AB$ и $V = BC$ и обозначить через UV. Если пути U, V соответственно эквивалентны путям U_1, V_1, то произведения их UV и U_1V_1 будут также эквивалентны. Умножение путей ассоциативно в том смысле, что если одно из произведений $U(VW)$ или $(UV)W$ определено, то другое также определено и оба представляют эквивалентные пути. Если подвижную точку M заставить пробегать какой-нибудь путь $U = \overset{\frown}{AB}$ в противоположном направлении, то получится обратный путь $U^{-1} = BA$, ведущий из точки B в точку A. Произведение пути AB на обратный ему путь BA будет замкнутым путем, эквивалентным нулевому пути, состоящему лишь из одной точки A.

Согласно определению, перемножать можно не любые два пути, а лишь такие, у которых конечная точка первого пути совпадает с начальной точкой второго. Этот недостаток исчезает, если рассматривать лишь замкнутые пути, выходящие из одной и той же начальной точки A. Любые два такие пути можно перемножить, в результате чего снова получится замкнутый путь с начальной точкой A. Кроме того, для каждого замкнутого пути с начальной точкой A обратный путь обладает теми же свойствами.

Условимся теперь эквивалентные пути считать различными представителями одного и того же «пути», лишь проведенного различными способами на поверхности, а неэквивалентные пути будем считать представителями существенно различных «путей». Приведенные выше замечания показывают, что в таком случае совокупность всех замкнутых путей (кавычки мы опускаем), выходящих из какой-либо точки A поверхности, будет являться группой относительно операции умножения путей. Единичным (нейтральным) элементом этой группы будет нулевой путь, а обратным элементом для данного пути будет служить этот же путь, только проходимый в обратном направлении.

Группа путей, вообще говоря, зависит не только от вида поверхности, но и от выбора начальной точки A. Однако если поверхность

не распадается на отдельные куски, т. е. если любые ее две точки могут быть соединены непрерывным путем, лежащим на поверхности, то группы путей, отвечающих различным точкам, будут изоморфными, и в этом случае можно говорить просто о группе путей поверхности S, не указывая точки A. Эта группа путей поверхности и называется ее *фундаментальной* группой.

Если поверхность S есть плоскость или сфера, то группа путей состоит лишь из единичного элемента, так как на плоскости и на сфере любой путь стягивается в точку. Однако уже на поверхности бесконечного кругового цилиндра, как мы видели, есть замкнутые пути, не стягиваемые в одну точку. Поскольку всякий замкнутый путь на цилиндре, выходящий из точки A, эквивалентен некоторой степени пути X (рис. 23), причем различные степени X между собой не эквивалентны, группа путей цилиндрической поверхности является бесконечной циклической группой. Можно доказать, что группа путей тора (рис. 22) состоит из путей вида U^mV^n (m, $n=0$, ±1, ±2, ...), причем $UV=VU$ и $U^mV^n=U^{m_1}V^{n_1}$ только при $m=m_1$, $n=n_1$ (напоминаем, что при рассмотрении группы путей равенство путей понимается в смысле их эквивалентности).

Важность группы путей объясняется следующим ее свойством. Допустим, что, кроме поверхности S, дана другая поверхность S_1, такая, что между точками поверхностей S и S_1 можно установить взаимно однозначное и непрерывное соответствие. Например, такое соответствие возможно, если поверхность S_1 получена из S посредством некоторой непрерывной деформации без разрывов и без склеиваний различных точек поверхности. Каждому пути на исходной поверхности S будет соответствовать путь на поверхности S_1. При этом эквивалентным путям будут соответствовать эквивалентные, произведению двух путей — произведение, так что группа путей на поверхности S_1 будет изоморфна группе путей на поверхности S. Иначе говоря, группа путей, рассматриваемая с абстрактной точки зрения, т. е. с точностью до изоморфизма, является инвариантом при всевозможных взаимно однозначных и непрерывных преобразованиях поверхности. Если группы путей для двух поверхностей различны, то соответствующие поверхности не могут быть переведены непрерывно одна в другую. Так, например, плоскость не может быть без склеиваний и разрывов деформирована в цилиндрическую поверхность, потому что группа путей плоскости состоит из единичного элемента, а группа путей цилиндра бесконечна.

Свойства фигур, остающиеся неизменными при взаимно однозначных и взаимно непрерывных преобразованиях, изучаются в особой математической дисциплине топологии, основные идеи которой были освещены в главе XVIII. Инварианты непрерывных преобразований называются *топологическими инвариантами*. Группа путей является одним из замечательнейших примеров топологических инвариантов. Ясно, что

группа путей может быть определена не только для поверхности, но и для любых множеств точек, лишь бы в этих множествах можно было говорить о путях и их деформациях.

Определяющие соотношения. Способы вычисления групп путей подробно изучены в топологии. При этом оказалось, что, как правило, указанные группы приходится определять при помощи особого способа, который часто применяется в теории групп для задания абстрактных групп, а не только для фундаментальных групп в топологии. Заключается он в следующем.

Пусть G — некоторая группа. Элементы $g_1, g_2, \ldots, g_n$ называются образующими элементами группы G, если всякий элемент g можно представить в виде

$$g = g_{i_1}^{\alpha_1} g_{i_2}^{\alpha_2} \ldots g_{i_k}^{\alpha_k},$$

где $i_1, i_2, \ldots, i_k$ — некоторые из чисел 1, 2, ..., n; не стоящие рядом номера i могут совпадать; число сомножителей k произвольно; показатели $\alpha_1, \alpha_2, \ldots, \alpha_k$ — не равные нулю положительные или отрицательные целые числа.

Чтобы знать группу G, достаточно, кроме образующих, знать еще, какие произведения предотавляют один и тот же элемент группы и какие представляют различные элементы. Таким образом, для задания группы нужно перечислить все равенства вида

$$g_{i_1}^{\alpha_1} g_{i_2}^{\alpha_2} \ldots g_{i_k}^{\alpha_k} = g_{j_1}^{\beta_1} g_{j_2}^{\beta_2} \ldots g_{j_l}^{\beta_l},$$

имеющие место в группе G. Так как таких равенств всегда бесконечное множество, то, вместо перечисления всех их, даются обычно лишь такие равенства, из которых все остальные вытекают в силу групповых аксиом. Эти равенства и называются определяющими соотношениями.

Ясно, что одна и та же группа может быть задана самыми различными определяющими соотношениями.

Рассмотрим, например, группу H с образующими a, b и соотношениями

$$a^2 = b^3, \quad ab = ba. \tag{10}$$

Положив $c = ab^{-1}$, будем иметь

$$a = bc, \quad a^2 = b^2c^2, \quad b^3 = b^2c^2, \quad b = c^2, \quad a = c^3.$$

Мы видим, что все элементы группы H можно выразить через один элемент c, причем

$$a = c^3, \quad b = c^2.$$

Так как из этих равенств соотношения (10) непосредственно вытекают, то для c никаких нетривиальных соотношений нет. Следовательно,

группа H является бесконечной циклической группой с образующим элементом c.

Если в группе можно выбрать такие образующие, которые не связаны никакими нетривиальными соотношениями, то группа называется *свободной*, а указанные образующие — свободными образующими. Если, например, группа имеет свободные образующие a, b, то всякий ее элемент *однозначно* записывается в форме

$$a^{\alpha_0}b^{\beta_1}a^{\alpha_1}b^{\beta_2}a^{\alpha_2}\ldots b^{\beta_k}a^{\alpha_k},$$

где $k=0, 1, 2, \ldots, n$, показатели α_0, β_1, α_1, $\ldots$, β_k, α_k представляют собой целые числа, положительные или отрицательные, отличные от нуля, кроме «крайних» α_0 и α_k, которые могут принимать и нулевые значения. Аналогичное утверждение справедливо и для свободных групп с большим числом образующих.

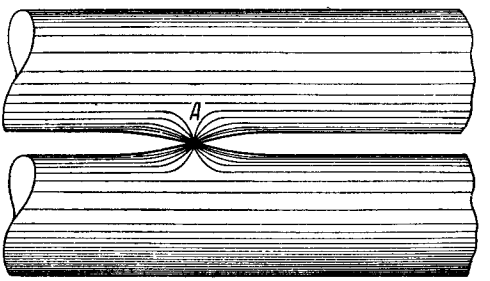


Рис. 24.

Если выписать образующие и определяющие соотношения для двух групп при условии, что рассматриваемые группы общих элементов не имеют, то, объединяя эти соотношения, мы получим новую группу, называемую свободным произведением данных.

Теория свободных групп, а также более общая теория свободных произведений, занимает в теории групп заметное место. С геометрической точки зрения свободное произведение групп H_1 и H_2 это группа путей такой фигуры, которую можно представить в виде суммы двух замкнутых фигур, склеенных лишь в одной точке и имеющих H_1 и H_2 своими группами путей. Мы знаем уже, что группа путей цилиндрической поверхности есть свободная группа с одним образующим. Из сделанного замечания следует, например, что группа путей поверхности, изображенной на рис. 24, есть свободная группа с двумя образующими.

Подобно тому, как определялась фундаментальная группа поверхности, можно ввести фундаментальную группу для пространственных тел, конечных или бесконечных.

Узлы и группы узлов. Как уже говорилось, с точки зрения топологии две поверхности считаются одинаковыми, если одну из них можно перевести в другую путем взаимно однозначного и взаимно непрерывного преобразования. Задача топологической классификации всех замкнутых поверхностей давно уже решена. Оказалось, что каждая замкнутая поверхность, расположенная в нашем обыкновенном пространстве, топологически эквивалентна либо сфере, либо сфере с несколькими ручками (рис. 25). Так, например, поверхность тора, изображенная на

рис. 22, может быть непрерывно деформирована в сферу с одной ручкой, поверхность куба — в поверхность сферы и т. п. Ввиду этого изучение фундаментальных групп для замкнутых поверхностей не очень интересно, так как замкнутые поверхности вполне классифицированы и без этих групп. Однако существуют очень простые задачи, где без

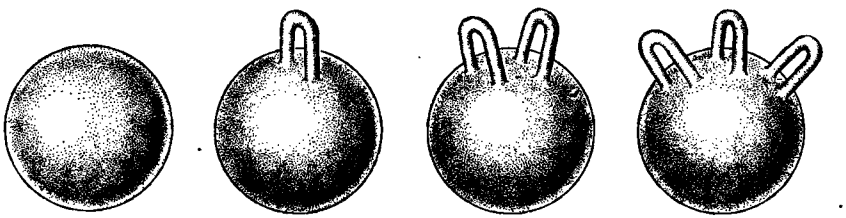


Рис. 25.

фундаментальных групп до сих пор почти ничего не удалось сделать. К числу их относится знаменитая проблема узлов.

Узлом мы будем называть замкнутую кривую, расположенную в обычном трехмерном пространстве. Это расположение может быть, как показывает рис. 26, весьма различным. Два узла называются эквивалентными, если один из них можно деформировать в другой непрерывным процессом, не разрывая кривой и не зацепляя ее саму за себя.

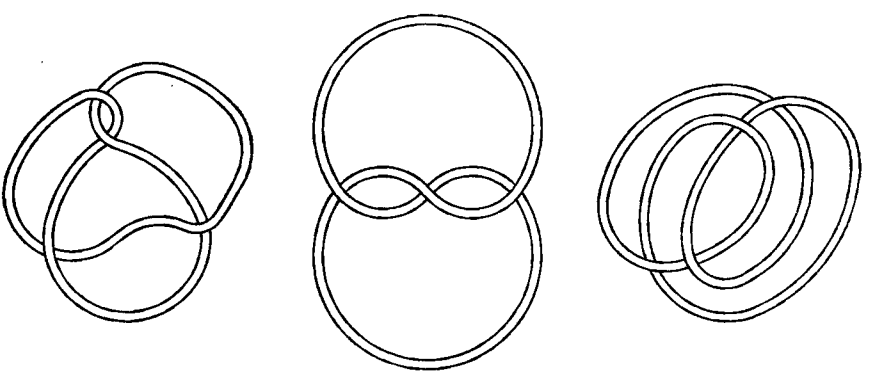


Рис. 26.

Сразу же возникают две задачи: 1) как узнать, эквивалентны или нет любые два узла, заданные своими плоскими чертежами; 2) как классифицировать все неэквивалентные узлы.

Обе задачи до сих пор остаются нерешенными, причем имеющиеся пока основные успехи в их частичном решении связаны с теорией групп.

Выбросим из пространства точки, принадлежащие данному узлу, и рассмотрим фундаментальную группу оставшегося множества точек. Эту группу и называют группой узла. Непосредственно очевидно, что

если узлы эквивалентны, то их группы изоморфны. Поэтому из неизоморфности групп узлов можно заключить о неэквивалентности самих узлов. Так, например, группа узла, приводящегося к окружности, есть циклическая группа, а группа узла, имеющего вид трилистника (рис. 27), есть более сложная группа. Последняя группа некоммутативна и, таким образом, неизоморфна группе окружности. Поэтому можно утверждать, что трилистниковый узел невозможно расправить в окружность, не разрубая его, — факт, очевидный экспериментально, но требующий для своего доказательства тонких математических соображений.

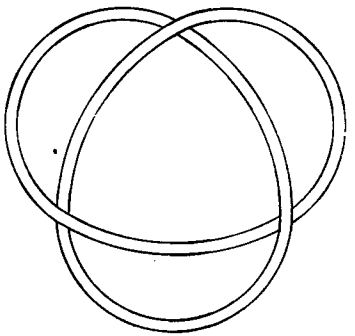


Рис. 27.

К сожалению, при рассмотрении групп узлов также встают не решенные до сих пор трудные задачи. Дело в том, что в топологии известны очень простые способы, как по заданному изображению узла найти образующие и определяющие соотношения группы узла. Но чтобы использовать группы для сравнения различных узлов, необходимо уметь решить, изоморфны или нет группы, заданные своими образующими и определяющими соотношениями, а решение этой задачи до сих пор не известно. Более того, советский математик П. С. Новиков недавно доказал замечательную теорему о том, что невозможно указать никакой единый регулярный способ (точнее — так называемый нормальный алгорифм), посредством которого можно было бы всегда решить, определяют ли две заданные системы определяющих соотношений для одних и тех же образующих одну и ту же группу или нет. Эта теорема заставляет невольно высказать сомнение и в существовании какого-либо единообразного общего способа для распознавания эквивалентности узлов, заданных своими плоскими изображениями.

§ 9. ПРЕДСТАВЛЕНИЯ И ХАРАКТЕРЫ ГРУПП

Общая теория групп по своим методам несколько напоминает элементарную геометрию: в обоих случаях в основу кладется определенная система аксиом, исходя из которой и строится все здание теории. Но пример аналитической геометрии показывает, насколько полезным для исследования геометрических проблем оказывается привлечение аналитических, числовых методов.

Приложением средств анализа и классической алгебры к теории групп является так называемая теория представлений групп. Как аналитическая геометрия дает методы решения не только геометрических задач при помощи анализа, но и, наоборот, бросает геометрический

свет на многие сложные проблемы анализа, так в еще большей степени теория представлений не только служит вспомогательным аппаратом для исследования свойств групп, но и, связывая воедино глубокие понятия и проблемы анализа и теории групп, позволяет для групповых фактов находить выражения в числовых соотношениях, а для аналитических зависимостей находить групповое истолкование. В настоящее время бóльшая часть важных приложений теории групп в физике связана именно с теорией представлений.

Представления групп матрицами. В линейной алгебре (см. главу XVI) уже рассматривалось действие умножения для матриц. Это действие ассоциативно, но, вообще говоря, некоммутативно. Неособенные квадратные матрицы данного порядка образуют группу относительно умножения. Действительно, произведение двух неособенных матриц есть снова неособенная матрица, роль нейтрального элемента играет единичная матрица и для каждой неособенной матрицы существует обратная, тоже неособенная.

Допустим, что дана некоторая группа G и каждому ее элементу g поставлена в соответствие определенная неособенная матрица комплексных чисел A_g порядка n, и притом так, что при перемножении элементов группы соответствующие им матрицы также перемножаются: $A_{gh} = A_g \cdot A_h$. Тогда говорят, что дано представление группы G матрицами степени n. Обычно слово «матрицами» пропускается и говорят просто о представлении степени n группы G. Представление степени n данной группы G есть просто гомоморфное отображение группы G в группу неособенных матриц степени n. Из общих свойств гомоморфных отображений следует, что при любом представлении нейтральный элемент группы G переходит в единичную матрицу, взаимно обратные элементы в G переходят во взаимно обратные матрицы.

Матрицы 1-го порядка — это отдельные комплексные числа. Поэтому представлением 1-й степени группы G является соответствие, при котором каждому элементу группы G отвечает комплексное число, причем произведению элементов группы отвечает произведение соответствующих комплексных чисел. Например, отображение, при котором четным подстановкам симметрической группы отвечает число 1, а нечетным число -1, будет представлением 1-й степени.

Поставив в соответствие каждому элементу группы G единичную матрицу E степени n, мы получим представление группы G, называемое ее единичным представлением степени n. Если группа G конечная и содержит неединичные элементы, то, кроме единичного представле-

группы G матрицами степени n. Выберем произвольную неособенную матрицу P той же степени n и положим $B_g = P^{-1}A_gP$. Соответствие $g \to B_g$ будет снова представлением группы G, так как

$$B_{gh} = P^{-1}A_{gh}P = P^{-1}A_gA_hP = P^{-1}A_gPP^{-1}A_hP = B_gB_h.$$

Представления, получающиеся этим способом из данного представления путем выбора различных матриц P, называются *эквивалентными* данному представлению. В теории представлений эквивалентные представления не считаются существенно различными, все представления рассматриваются обычно лишь с точностью до эквивалентности.

Другим способом нахождения новых представлений является прямое сложение представлений, заключающееся в следующем: пусть $g \to A_g$, $g \to B_g$ — какие-либо представления группы G матрицами соответственно степеней m и n. Рассмотрим отображение

$$g \to \begin{pmatrix} A_g & 0 \\ 0 & B_g \end{pmatrix}.$$

Учитывая правила перемножения матриц (см. главу XVI), имеем

$$gh \to \begin{pmatrix} A_{gh} & 0 \\ 0 & B_{gh} \end{pmatrix} = \begin{pmatrix} A_gA_h & 0 \\ 0 & B_gB_h \end{pmatrix} = \begin{pmatrix} A_g & 0 \\ 0 & B_g \end{pmatrix}\begin{pmatrix} A_h & 0 \\ 0 & B_h \end{pmatrix},$$

т. е. указанное отображение снова является представлением группы G. Оно называется суммой двух данных представлений и обозначается $A_g + B_g$. Если слагаемые переставить, то получится иное представление

$$g \to \begin{pmatrix} B_g & 0 \\ 0 & A_g \end{pmatrix},$$

которое, однако, эквивалентно старому. Поэтому, если эквивалентные представления не различать, то сложение представлений будет коммутативной операцией. Легко видеть, что при том же условии сложение будет и ассоциативной операцией. Имея некоторый запас представлений A_g, B_g, C_g, ... группы G, можно путем их сложения получать представления все более высоких степеней: $A_g + B_g + C_g$, $A_g + A_g + A_g + A_g$ и т. д.

Например, числа 1, -1, i, $-i$ относительно умножения образуют группу. Ставя каждому числу этой группы в соответствие само это число, мы получим представление 1-й степени. В качестве второго представления можно взять отображение $1 \to 1$, $-1 \to -1$, $i \to -i$, $-i \to i$. Суммой этих представлений будет отображение

$$1 \to \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix},\ -1 \to \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix},\ i \to \begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix},\ -i \to \begin{pmatrix} -i & 0 \\ 0 & i \end{pmatrix}.$$

Преобразуя его при помощи матрицы $P=\begin{pmatrix}1 & i\\ 1 & -i\end{pmatrix}$, получим эквивалентное представление

$$1\to\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix},\quad -1\to\begin{pmatrix}-1 & 0\\ 0 & -1\end{pmatrix},\quad i\to\begin{pmatrix}0 & 1\\ -1 & 0\end{pmatrix},\quad -i\to\begin{pmatrix}0 & -1\\ 1 & 0\end{pmatrix}.$$

Интересно отметить, что все матрицы этого представления действительны.

Допустим, что все матрицы некоторого представления степени n группы G имеют вид

$$g\to A_g=\begin{pmatrix}B_g & C_g\\ 0 & D_g\end{pmatrix},$$

где B_g, D_g — квадратные матрицы, а левый нижний угол A_g целиком заполнен нулями. Перемножая матрицы A_g и A_h, получим

$$A_{gh}=A_gA_h=\begin{pmatrix}B_gB_h & B_gC_h+C_gD_h\\ 0 & D_gD_h\end{pmatrix},$$

т. е. $B_{gh}=B_gB_h$, $D_{gh}=D_gD_h$. Это показывает, что отображения $g\to B_g$ и $g\to D_g$ будут также представлениями группы G, но только более низкой степени. Представление A_g называется ступенчатым представлением группы G, а всякое эквивалентное ему представление называется *приводимым*. Представление же, не эквивалентное никакому ступенчатому, называется *неприводимым*.

Если во всех матрицах A_g не только левый нижний, но и правый верхний угол C_g заполнен нулями, то представление A_g называется *распадающимся* в сумму представлений B_g, D_g. Представление, эквивалентное сумме неприводимых представлений, называется *вполне приводимым*.

В теории групп доказывается, что всякое представление конечной группы вполне приводимо[1]. Отсюда следует, что для нахождения всех представлений конечной группы достаточно знать ее неприводимые представления, так как все остальные эквивалентны различным суммам неприводимых.

Практическое вычисление неприводимых представлений какой-либо конечной группы является, обычно, довольно сложной задачей, которая в явной форме решена лишь для отдельных классов конечных групп, например, для коммутативных групп, для симметрических групп и некоторых других, хотя с теоретической точки зрения свойства представлений конечных групп изучены довольно подробно.

---

[1] Напомним, что мы рассматриваем представления групп матрицами, элементами которых могут являться произвольные комплексные числа.

Для каждой конечной группы вводится особое «регулярное» представление, которое строится следующим образом. Пусть $g_1, g_2, \ldots, g_n$ — элементы заданной группы G, перенумерованные в произвольном порядке, и пусть

$$g_i g_k = g_{ik} \quad (i, k = 1, 2, \ldots, n).$$

Выбирая какое-либо фиксированное значение для k, составим матрицу степени n, у которой в i-й строке стоит 1 на i_k-м месте, а остальные места заняты нулями ($i = 1, 2, \ldots, n$), и введем для нее обозначение R_{g_k}. Соответствие $g_k \to R_{g_k}$ ($k = 1, 2, \ldots, n$) называется *регулярным представлением* группы G. То, что это действительно представление, доказывается путем простых вычислений.

Можно доказать также, что, изменяя нумерацию элементов группы, мы перейдем к эквивалентному представлению, и, следовательно, с точностью до эквивалентности каждая конечная группа имеет лишь одно регулярное представление.

Формулируем вкратце основные теоремы теории представлений конечных групп. Число различных (неэквивалентных) неприводимых представлений конечной группы конечно и равно числу классов сопряженных (см. стр. 280) элементов этой группы. Степени неприводимых представлений обязательно являются делителями порядка группы, причем регулярное представление эквивалентно сумме всех неэквивалентных неприводимых представлений, в которой каждое неприводимое слагаемое повторяется столько раз, какова его степень.

Отсюда вытекает следующее интересное соотношение между порядком конечной группы и степенями неприводимых представлений.

Обозначим число элементов группы G через n, число классов сопряженных элементов через k и степени неприводимых представлений G соответственно через $n_1, n_2, \ldots, n_k$. Из построения регулярного представления видно, что его степень равна n. Поскольку, кроме того, регулярное представление эквивалентно сумме n_1 представлений, эквивалентных первому неприводимому представлению, плюс n_2 представлений, эквивалентных второму, и т. д., и при сложении представлений их степени складываются, должно иметь место равенство

$$n = n_1^2 + n_2^2 + \ldots + n_k^2. \tag{11}$$

Поставив каждому элементу группы в соответствие число 1, мы получим тривиальное неприводимое представление степени 1, которым обладает каждая группа. Понимая под n_1 в формуле (11) степень именно этого единичного представления, можно формулу (11) переписать в равносильной форме

$$n = 1 + n_2^2 + \ldots + n_k^2,$$

где $n_2, \ldots, n_k$ означают теперь степени нетривиальных неприводимых представлений.

Пользуясь тем, что $n_2, \ldots, n_k$ должны быть делителями числа n и зная k, иногда можно только из одного равенства (11) уже найти $n_2, \ldots, n_k$. Например, симметрическая группа S_3 перестановок трех элементов имеет три класса сопряженных перестановок: (1); (12), (13), (23); (123), (132). При $n=6$, $k=3$ равенство (11) допускает единственную систему решений: $6=1^2+1^2+2^2$. Поэтому S_3 имеет два различных представления 1-й степени и одно неприводимое представление 2-й степени.

Другим примером могут служить конечные абелевы группы. Здесь каждый элемент составляет отдельный класс. Поэтому $k=n$, и из формулы (11) следует, что $n_1=n_2=\ldots=n_k=1$, т. е. все неприводимые представления таких групп имеют первую степень и число их равно порядку группы.

Неприводимые представления абелевых групп иначе называются их характерами. Для всякого представления неабелевой группы характером называется набор так называемых следов (т. е. сумм диагональных элементов) матриц, образующих представление. Характеры конечных групп обладают замечательными свойствами и соотношениями. Исследование представлений и характеров групп обогатило теорию групп интересными общими результатами, нашедшими обширные приложения и в современной теоретической физике.

§ 10. ОБЩАЯ ТЕОРИЯ ГРУПП

Мы уже отмечали, что в течение почти всего прошлого века теория групп развивалась преимущественно как теория групп преобразований. Однако постепенно выяснилось, что основным является изучение именно групп, как таковых, а изучение групп преобразований может быть сведено к изучению абстрактных групп и их подгрупп.

Переход от теории групп преобразований к теории абстрактных групп совершился вначале в теории конечных групп, но быстрое развитие теории групп Ли, а также проникновение теории групп в топологию сделали необходимым создание общей теории групп, в которой конечные группы рассматривались бы лишь как некоторый частный случай.

Первым руководством по теории групп, в котором со всей отчетливостью была проведена эта точка зрения, явилась книга О. Ю. Шмидта, вышедшая в Киеве в 1916 г. В 20-х годах Шмидт получил также важную теорему, относящуюся к теории бесконечных групп, которая стала отправным пунктом исследований для ряда других советских алгебраистов. Благодаря деятельности О. Ю. Шмидта и П. С. Александрова, много сделавших для популяризации идей современной алгебры, в Москве образовалась крупная школа теории групп, руководителем которой вскоре стал их ученик А. Г. Курош.

Широкую известность приобрела, в частности, доказанная им теорема о том, что всякая подгруппа свободного произведения сама является свободным произведением подгрупп, изоморфных соответствующим подгруппам сомножителей, и, может быть, еще отдельной свободной подгруппы. Позже им была опубликована монография по теории групп, в которой впервые был систематически изложен богатый фактический материал, полученный в области общей теории групп. Эта монография в настоящее время является наиболее полным в мировой литературе курсом общей теории групп, пользующимся широкой международной известностью.

Вслед за алгебраистами Москвы общей теорией групп стали заниматься алгебраисты Ленинграда и других городов, внесшие большой вклад в ее развитие. Исследования по теории групп, ведущиеся в настоящее время в СССР, охватывают все ее существенные разделы, а полученные советскими математиками результаты уже неоднократно оказывали решающее влияние на развитие теории групп.

§ 11. ГИПЕРКОМПЛЕКСНЫЕ ЧИСЛА

При решении практических задач алгебраическим методом обычно в простейших случаях приходят к одному или нескольким уравнениям, из которых находят значения неизвестных величин. Неизвестные при этом являются количественными характеристиками изучаемых объектов; уравнения же составляются при помощи анализа реальных отношений, существующих между объектами.

Так обстоит дело в случаях, когда речь идет о простейших величинах, подобных массе, объему или расстоянию, для количественной характеристики которых достаточно одного числа. Однако в конкретных задачах встречаются не только объекты, характеризующиеся одним числом. Напротив, с развитием техники все большее значение приобретают объекты более сложной природы, для характеристики которых необходимо несколько и даже бесконечно много чисел. Уже такие важные физические величины, как сила, скорость, ускорение, характеризуются направленным отрезком и требуют для своего задания трех чисел. Далее известно, что положение точки в пространстве характеризуется тремя числами, положение плоскости — также тремя, положение прямой — четырьмя, а положение твердого тела — даже шестью числами. Поэтому когда приходится при помощи алгебры решать задачи, касающиеся более сложных объектов, то получаются уравнения с большим числом неизвестных, разобраться в которых часто оказывается труднее, чем решить задачу непосредственно, пользуясь ее геометрическими или физическими особенностями. Отсюда естественно возникла мысль попытаться характеризовать более сложные объекты не системами

обыкновенных чисел, а какими-нибудь более сложными обобщенными числами, над которыми можно было бы совершать операции, похожие на обычные арифметические действия. Эта постановка вопроса была тем более естественна, что история науки показывала не неизменность понятия числа, а его изменчивость, постепенное обогащение совокупности чисел от чисел натуральных к числам дробным, затем к относительным числам, действительным (рациональным и иррациональным) и, наконец, к комплексным.

Комплексные числа. Из главы IV (том 1) читателю уже известны основные свойства комплексных чисел и простейшие их приложения. Сейчас нас будет интересовать только обоснование понятия комплексного числа. Напомним, как обычно определяется само понятие комплексного числа. Сначала рассматривают только обыкновенные действительные числа и замечают, что квадратный корень из отрицательных чисел не имеет смысла, поскольку квадрат каждого действительного числа есть число положительное или нуль. Далее указывают, что неотложные потребности науки заставили математиков рассматривать и выражения вида $a+b\sqrt{-1}$ как особого рода числа, которые они, в отличие от обыкновенных, действительных чисел, стали называть мнимыми. Если считать, что эти мнимые числа подчиняются тем же правилам арифметических действий, что и обычные числа, то все корни квадратные из отрицательных чисел могут быть выражены через величину $i=\sqrt{-1}$, а результат арифметических действий, произведенных в любом конечном числе над действительными и мнимыми числами, может быть всегда представлен в виде $a+bi$, где a и b — действительные числа.

Ясно, что такое определение мнимых чисел в высшей степени противоречит здравому смыслу, так как сначала утверждается, что выражения $\sqrt{-1}$, $\sqrt{-2}$ и т. п. смысла не имеют, а затем предлагается эти не имеющие смысла выражения называть мнимыми числами. Это обстоятельство вызвало большие сомнения математиков XVII и XVIII вв. в законности пользования комплексными числами. Однако эти сомнения рассеялись в начале XIX в., когда нашли геометрическое изображение комплексных чисел точками на плоскости. Строгое чисто арифметическое обоснование теории комплексных чисел было дано немного позже венгерским математиком Бóйаи и английским — Гамильтоном. Это обоснование заключается в следующем.

Вместо чисел $a+bi$ будем говорить просто о парах действительных чисел (a, b). Две пары условимся называть равными, если равны соответственно их первые и вторые члены, т. е. $(a, b)=(c, d)$ тогда и только тогда, когда $a=c$ и $b=d$. Сложение и умножение пар определим формулами

$$(a, b)+(c, d)=(a+c,\ b+d);\quad (a, b)\cdot(c, d)=(ac-bd,\ ad+bc).$$

Например, имеем

$$(2,\ 3)+(1,\ -2)=(3,\ 1),\quad (2,\ 3)(1,\ -2)=(8,\ -1),$$
$$(3,\ 0)+(2,\ 0)=(5,\ 0),\quad (3,\ 0)(2,\ 0)=(6,\ 0).$$

Эти примеры показывают, в частности, что арифметические действия над парами, у которых на втором месте стоит 0, приводятся к тем же действиям над их первыми членами, в силу чего такие пары можно обозначить просто их первым числом. Вводя еще обозначение i для пары (0, 1), будем иметь

$$(a,\ b)=a(1,\ 0)+b(0,\ 1)=a+bi,$$
$$i^2=(0,\ 1)(0,\ 1)=(-1,\ 0)=-1,$$

т. е. будем иметь обычные обозначения комплексных чисел.

Итак, с изложенной точки зрения комплексные числа являются парами обычных действительных чисел и действия с комплексными числами — лишь особого рода действия с парами действительных чисел.

Гиперкомплексные числа. Многообразное и успешное применение комплексных чисел побудило математиков уже в первые десятилетия XIX в. задуматься над вопросом, нельзя ли подобно тому, как комплексные числа строятся в виде пар действительных чисел, построить высшие комплексные числа, изображающиеся тройками, четверками и т. д. действительных чисел. Начиная с середины прошлого века было исследовано много различных частных систем таких высших комплексных или гиперкомплексных чисел, а в конце прошлого и первой половине текущего столетия была разработана общая теория гиперкомплексных чисел, нашедшая ряд важных приложений в смежных областях математики и физики.

Итак, назовем гиперкомплексным числом ранга n число, изображающееся совокупностью n действительных чисел $(a_1, a_2, \ldots, a_n)$, которые пока условно будем называть координатами этого гиперкомплексного числа. Гиперкомплексные числа $(a_1, a_2, \ldots, a_n)$ и $(b_1, b_2, \ldots, b_n)$ будем называть равными, если равны их соответствующие координаты, т. е. если $a_1=b_1$, $a_2=b_2$, $\ldots$, $a_n=b_n$. Действие сложения определим естественной формулой

$$(a_1, a_2, \ldots, a_n)+(b_1, b_2, \ldots, b_n)=(a_1+b_1, a_2+b_2, \ldots, a_n+b_n),$$

аналогичной формуле сложения для комплексных чисел.

Так же естественно вводится операция умножения гиперкомплексного числа на действительное: по определению считается, что

$$a(a_1, a_2, \ldots, a_n)=(aa_1, aa_2, \ldots, aa_n).$$

Сверх того, должно быть определено действие умножения двух гиперкомплексных чисел друг на друга, причем результат этого действия должен являться гиперкомплексным числом.

Распространить на общий случай определение умножения обыкновенных комплексных чисел трудно. Оно может быть осуществлено различными путями, и при этом будут получаться различные системы гиперкомплексных чисел. Поэтому прежде всего следует уяснить, что должно быть достигнуто таким определением. Несомненно желательно, чтобы определяемые нами действия над гиперкомплексными числами были похожи по своим свойствам на обычные действия с действительными числами. Каковы же свойства этих обычных действий?

Рассматривая внимательно свойства чисел и операций над ними, наиболее часто используемые в алгебре, легко заметить, что они сводятся к следующим.

1. Для любых двух чисел однозначно определена их сумма.
2. Для любых двух чисел однозначно определено их произведение.
3. Существует число нуль со свойством $a+0=a$ для любого a.
4. Для каждого числа a существует противоположное число x, удовлетворяющее равенству $a+x=0$.
5. Сложение переместительно (коммутативно)

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

6. Сложение обладает сочетательным (ассоциативным) свойством

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

7. Умножение переместительно

$$ab=ba.$$

8. Умножение сочетательно

$$(ab)\cdot c=a\cdot(bc).$$

9. Умножение распределительно (дистрибутивно)

$$a(b+c)=ab+ac, \quad (b+c)a=ba+ca.$$

10. Для каждого a и каждого $b\neq 0$ существует единственное число x, удовлетворяющее равенству $bx=a$.

Свойства 1—10 были выделены в результате тщательного анализа; развитие математики в последнее столетие показало их большую важность. В настоящее время всякая система величин, удовлетворяющая условиям 1—10, называется *полем*. Например, полями являются: совокупность всех рациональных чисел, совокупность всех действительных чисел, совокупность всех комплексных чисел, так как числа каждой из этих совокупностей можно складывать, перемножать, получая числа из той же совокупности, и эти действия обладают свойствами 1—10. Кроме этих трех наиболее важных полей, можно указать бесчисленное множество других полей, образованных числами. Однако наряду с полями, образованными числами, большой интерес представляют

и поля, образованные величинами иной природы. Например, еще в средней школе нас обучают действиям с так называемыми алгебраическими дробями, т. е. дробями, у которых числитель и знаменатель являются многочленами относительно некоторых букв. Алгебраические дроби можно складывать, вычитать, перемножать, делить, и эти действия обладают свойствами 1—10. Следовательно, алгебраические дроби образуют систему объектов, являющуюся полем. Можно привести много других примеров полей, составленных величинами все более и более сложной природы. Ввиду важности свойств 1—10, определяющих поле, первоначально была поставлена задача найти такое действие умножения гиперкомплексных чисел, чтобы гиперкомплексные числа образовали поле. В случае успеха тем самым были бы получены новые, более общие комплексные числа. Однако уже в самом начале прошлого века обнаружили, что это возможно лишь для гиперкомплексных чисел ранга 2, причем получаются только обыкновенные комплексные числа. Этот результат показал, что комплексные числа занимают совершенно особое место и что получить расширение числовой системы за пределы комплексных чисел нельзя, если непременно требовать, чтобы выполнялись все свойства 1—10. Поэтому при дальнейших попытках построения высших числовых систем необходимо было отказаться от одного или нескольких свойств 1—10.

Кватернионы. Исторически первой гиперкомплексной системой, рассмотренной в математике, является система кватернионов, т. е. «четверных чисел», введенная английским математиком и механиком Гамильтоном в середине прошлого столетия. Эта система удовлетворяет всем требованиям 1—10, кроме 7 (коммутативность умножения).

Кватернионы описываются следующим образом. Введем для четверок $(1, 0, 0, 0)$, $(0, 1, 0, 0)$, $(0, 0, 1, 0)$, $(0, 0, 0, 1)$ сокращенные обозначения 1, i, j, k. Тогда в силу равенства

$$(a, b, c, d) = a(1, 0, 0, 0) + b(0, 1, 0, 0) + c(0, 0, 1, 0) + d(0, 0, 0, 1)$$

каждый кватернион можно однозначно представить в виде

$$(a, b, c, d) = a \cdot 1 + b \cdot i + c \cdot j + d \cdot k.$$

Кватернион 1 будем считать единицей строящейся системы величин, т. е. будем считать, что $1 \cdot \alpha = \alpha \cdot 1 = \alpha$ для любого кватерниона α. Далее положим по определению: $i^2 = j^2 = k^2 = -1$;

$$ij = -ji = k,$$
$$ik = -ki = -j,$$
$$jk = -kj = i.$$

Эту «таблицу умножения» кватернионов легко запомнить при помощи рис. 28, на котором точками i, j, k на окружности изображены последовательные кватернионы i, j, k. Произведение двух рядом располо-

женных кватернионов равно третьему, если движение от первого множителя ко второму происходит на рисунке по часовой стрелке, и равно третьему со знаком минус, если движение происходит против часовой стрелки. Зная таблицу умножения кватернионов i, j, k, умножение произвольных кватернионов производят, пользуясь распределительным законом 9. Именно:

$$(a\cdot 1+b\cdot i+c\cdot j+d\cdot k)(a_1\cdot 1+b_1\cdot i+c_1\cdot j+d_1\cdot k)=$$
$$=aa_1\cdot 1+ab_1\cdot i+ac_1\cdot j+ad_1\cdot k+ba_1\cdot i+bb_1\cdot ii+bc_1\cdot ij+bd_1\cdot ik+$$
$$+ca_1\cdot j+cb_1\cdot ji+cc_1\cdot jj+cd_1\cdot jk+da_1\cdot k+db_1\cdot ki+dc_1\cdot kj+dd_1\cdot kk=$$
$$=(aa_1-bb_1-cc_1-dd_1)\cdot 1+(ab_1+ba_1+cd_1-dc_1)\cdot i+$$
$$+(ac_1+ca_1-bd_1+db_1)\cdot j+(ad_1+da_1+bc_1-cb_1)\cdot k.$$

Множитель 1 в первом члене кватерниона обычно опускают и вместо $a\cdot 1$ пишут a. Равенства $ij=-ji$, $ik=-ki$, $jk=-kj$ показывают, что умножение кватернионов не перестановочно. Множимое и множитель здесь не равноправны. Поэтому при вычислениях с кватернионами необходимо тщательно следить за порядком сомножителей. В остальном действия с кватернионами не отличаются какой-либо трудностью. В частности, сочетательный закон 8 при умножении кватернионов имеет место. Он легко проверяется для базисных кватернионов 1, i, j, k при помощи таблицы умножения; переход же к общему случаю очевиден.

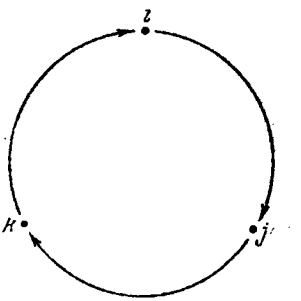


Рис. 28.

Число a кватерниона $a+bi+cj+dk$ называется его действительной или скалярной частью, а сумма $bi+cj+dk$ — его векторной частью. Кватернионы $a+bi+cj+dk$ и $a-bi-cj-dk$, отличающиеся лишь знаком векторной части, называются сопряженными. Очевидно, сумма двух сопряженных кватернионов есть число действительное. Кроме того, перемножая сопряженные кватернионы по приведенной выше формуле, получим

$$(a+bi+cj+dk)(a-bi-cj-dk)=a^2+b^2+c^2+d^2, \qquad (12)$$

т. е. действительным числом является и произведение сопряженных кватернионов.

Сумма квадратов коэффициентов $a^2+b^2+c^2+d^2$ кватерниона $a+bi+cj+dk$ называется его нормой. Поскольку квадрат любого действительного числа есть число неотрицательное, то норма каждого кватерниона есть также число неотрицательное, равное нулю только для нулевого кватерниона.

Формула (12) показывает, что произведение какого-либо кватерниона на сопряженный кватернион равно норме данного кватерниона.

Условимся звездочкой обозначать кватернион, сопряженный данному. Тогда непосредственное перемножение показывает справедливость следующей формулы:

$$(\alpha\beta)^* = \beta^*\alpha^*.$$

Отсюда вытекает интересное следствие: норма произведения кватернионов равна произведению норм сомножителей. В самом деле, на основании предыдущего имеем

$$\text{норма } (\alpha\beta) = (\alpha\beta)(\alpha\beta)^* = \alpha\beta\beta^*\alpha^* = (\alpha\alpha^*)(\beta\beta^*) = \text{норма } \alpha \cdot \text{норма } \beta.$$

Свойства нормы позволяют весьма просто решить и вопрос о делении кватернионов. Пусть $\alpha = a + bi + cj + dk$ — произвольный ненулевой кватернион. Тогда

$$(a + bi + cj + dk)\,\frac{1}{a^2+b^2+c^2+d^2}(a - bi - cj - dk) =$$

$$= \frac{1}{a^2+b^2+c^2+d^2}(a^2 + b^2 + c^2 + d^2) = 1,$$

т. е. кватернион

$$\frac{1}{a^2+b^2+c^2+d^2}(a - bi - cj - dk) = \alpha^{-1}$$

является обратным для заданного кватерниона α.

Умея находить обратный кватернион, легко найти и частные двух кватернионов. Действительно, пусть даны два кватерниона α, β, из которых первый отличен от нуля. Тогда частными от деления β на α мы должны назвать решения уравнений

$$\alpha x = \beta, \qquad y\alpha = \beta.$$

Умножая обе части первого уравнения на обратный кватернион α^{-1} слева, получим

$$x = \alpha^{-1}\beta.$$

Умножая обе части второго уравнения на α^{-1} справа, будем иметь

$$y = \beta\alpha^{-1}.$$

Так как произведения $\alpha^{-1}\beta$ и $\beta\alpha^{-1}$ в общем случае различны, то для кватернионов приходится различать два деления — правое и левое; оба они всегда выполнимы, за исключением, конечно, деления на нуль.

Алгебра векторов. Хотя действия с кватернионами во многом сходны с действиями над комплексными числами, все же отсутствие переместительного закона для умножения делает свойства кватернионов глубоко отличными от свойств чисел. Например, из алгебры комплексных чисел хорошо известно, что квадратное уравнение имеет два корня. Если же мы будем рассматривать хотя бы квадратное уравнение

$$x^2 + 1 = 0$$

в области кватернионов, то найдем сразу 6 его корней: $\pm i$, $\pm j$, $\pm k$, а более точный анализ показывает, что имеется еще бесчисленное множество других решений. Это обстоятельство сильно затрудняет использование кватернионов в математике, и, несмотря на многочисленные попытки Гамильтона и других математиков ввести кватернионы в различные отделы математики и физики, роль кватернионов остается до сих пор в математике относительно скромной и ни в какой мере несравнимой с ролью комплексных чисел.

Однако кватернионы дали толчок развитию векторной алгебры, являющейся незаменимым средством в современной технике и физике. Дело в том, что в механике и физике существенную роль играют понятия скорости, ускорения, силы и т. д., для характеристики которых нужны три числа. Выше мы видели, что каждый кватернион может быть рассматриваем как совокупность действительного числа a и векторной части $bi+cj+dk$. Поскольку векторная часть кватерниона определяется тремя числами, для характеристики важнейших физических величин достаточно уже векторных частей кватернионов.

Геометрически векторную часть $bi+cj+dk$ кватерниона $a+bi+cj+dk$ принято изображать вектором, выходящим из начала прямоугольной декартовой системы координат, проекции которого на оси координат соответственно равны числам b, c, d. Поэтому произвольный кватернион геометрически можно представлять как совокупность числа и вектора в пространстве. Посмотрим, какое истолкование при этом получат действия с кватернионами.

Возьмем два векторных кватерниона $xi+yj+zk$ и $x_1i+y_1j+z_1k$, скалярная часть которых равна нулю. Геометрически они изображаются векторами, выходящими из начала координат. Сумма этих кватернионов будет снова векторным кватернионом $(x+x_1)i+(y+y_1)j+(z+z_1)k$. Легко видеть, что вектор, изображающий эту сумму, будет диагональю параллелограма, построенного на первых двух векторах. Таким образом, сложению векторных кватернионов отвечает хорошо известная операция сложения векторов по правилу параллелограма. Аналогично, если умножить векторный кватернион на какое-либо действительное число, то изображающий кватернион вектор в результате также умножится на это число.

С иным положением мы сталкиваемся при перемножении кватернионов. Действительно,

$$(xi+yj+zk)(x_1i+y_1j+z_1k)=$$
$$=-xx_1-yy_1-zz_1+(yz_1-y_1z)i+(zx_1-z_1x)j+(xy_1-x_1y)k,$$

т. е., перемножая два векторных кватерниона, мы получаем полный кватернион, имеющий скалярную часть и векторную часть.

Скалярная часть произведения векторных кватернионов, взятая с обратным знаком, называется скалярным произведением векторов,

изображающих данные кватернионы, а вектор, изображающий векторную часть произведения, — векторным произведением указанных векторов. Скалярное произведение векторов α и β обозначается обычно через $(\alpha\ \beta)$ или просто $\alpha\beta$, а векторное произведение тех же векторов — через $[\alpha\beta]$. Пусть $\boldsymbol{i}$, $\boldsymbol{j}$, $\boldsymbol{k}$ — векторы, отвечающие кватернионам i, j, k, т. е. векторы единичной длины, отложенные вдоль соответственных осей координат. Согласно определению, если $\alpha = x\boldsymbol{i} + y\boldsymbol{j} + z\boldsymbol{k}$, $\beta = x_1\boldsymbol{i} + y_1\boldsymbol{j} + z_1\boldsymbol{k}$,

то

$$(\alpha\beta) = xx_1 + yy_1 + zz_1,\ [\alpha\beta] = (yz_1 - y_1z)\,\boldsymbol{i} + (zx_1 - z_1x)\,\boldsymbol{j} + (xy_1 - x_1y)\boldsymbol{k}.$$

При помощи последних формул легко дать и геометрическое истолкование скалярному и векторному произведениям векторов. Оказывается, скалярное произведение двух векторов равно произведению длин этих векторов на косинус угла между ними, а векторное произведение двух векторов есть вектор, по длине равный площади параллелограма, построенного на данных векторах, и направленный перпендикулярно площадке указанного параллелограма в ту сторону, откуда вращение от первого данного вектора ко второму кажется совершающимся в ту же сторону, что и вращение от оси Ox к оси Oy, если смотреть со стороны оси Oz.

В настоящее время в механике и физике не употребляются, как правило, действия с кватернионами, а вместо них рассматриваются лишь действия над векторами, причем эти действия определяются чисто геометрическим способом, следуя сформулированным только что результатам.

В заключение укажем одну задачу из механики, решаемую при помощи кватернионов особенно красиво. Решение ее собственно и послужило одной из причин открытия кватернионов.

Пусть твердое тело сначала поворачивается на некоторый угол φ в заданном направлении около определенной оси OA, проходящей через заданную точку O, после чего поворачивается на угол φ_1 около другой оси OB, проходящей через ту же точку. Спрашивается, около какой оси и на какой угол следует повернуть тело, чтобы оно из первого положения сразу перешло в третье? Это известная задача механики о сложении конечных поворотов. Правда, она может быть решена средствами обычной аналитической геометрии, что и было сделано еще Эйлером в XVIII в. Однако гораздо более прозрачную форму имеет ее решение при помощи кватернионов.

Пусть $\xi = x\boldsymbol{i} + y\boldsymbol{j} + z\boldsymbol{k}$ и $\alpha = a + b\boldsymbol{i} + c\boldsymbol{j} + d\boldsymbol{k}$ — два кватерниона, из которых первый мы будем считать переменным, а второй заданным. Выражение $\alpha^{-1}\xi\alpha$, как легко проверить вычислением, будет векторным кватернионом. Если теперь кватернионы ξ, $\alpha^{-1}\xi\alpha$ и векторную часть кватерниона α изобразить векторами $\vec{\xi}$, $\vec{\xi}_1$, $\vec{\alpha}$, то окажется, что век-

тор $\vec{\xi}_1$ геометрически получается из вектора $\vec{\xi}$ поворотом вокруг оси, проходящей через вектор $\vec{\alpha}$, на угол φ, определяемый формулой $\cos\frac{\varphi}{2}=\frac{a}{\sqrt{a^2+b^2+c^2+d^2}}$. Поэтому можно считать, что кватернион $\alpha=a+bi+cj+dk$ изображает поворот пространства на угол φ вокруг оси $\vec{\alpha}=bi+cj+dk$.

Обратно, зная ось поворота и угол φ, можно искать кватернион, изображающий этот поворот. Таких кватернионов оказывается бесконечное множество, но все они отличаются друг от друга лишь численным множителем.

Рассмотрим теперь еще один поворот на угол φ_1 вокруг некоторой оси $\vec{\beta}=b_1i+c_1j+d_1k$. Пусть этот поворот изображается кватернионом $\beta=a_1+b_1i+c_1j+d_1k$. Под воздействием первого поворота произвольный вектор $\vec{\xi}=xi+yj+zk$ перейдет в вектор $\alpha^{-1}\vec{\xi}\alpha$, а под воздействием второго поворота этот последний вектор перейдет в $\beta^{-1}(\alpha^{-1}\vec{\xi}\alpha)\beta$. На основании закона ассоциативности последний результат можно представить в форме

$$\beta^{-1}(\alpha^{-1}\vec{\xi}\alpha)\beta=(\alpha\beta)^{-1}\vec{\xi}\alpha\beta.$$

Поскольку умножение вектора, т. е. векторного кватерниона $\vec{\xi}$ на кватернион $(\alpha\beta)^{-1}$ слева и кватернион $\alpha\beta$ справа равносильно повороту этого вектора на соответствующий угол вокруг соответствующей оси, мы приходим к выводу, что результат двух последовательных поворотов, характеризующихся кватернионами α и β, есть поворот, характеризующийся произведением $\alpha\beta$. Иными словами, сложению поворотов отвечает перемножение соответствующих им кватернионов.

Помимо геометрических и физических приложений, кватернионы нашли замечательные приложения и в теории чисел. Из последующих работ в этой области следует отметить в особенности работы Ю. В. Линника.

§ 12. АССОЦИАТИВНЫЕ АЛГЕБРЫ

Общее определение алгебр (гиперкомплексных систем). Гиперкомплексные числа определялись как величины, для задания которых необходимо несколько действительных чисел, причем для определенности гиперкомплексные числа рассматривались просто как системы действительных чисел. Однако такая точка зрения слишком узка, и для теоретических исследований постепенно стали применять следующее более общее определение.

Некоторая система [величин] S называется алгеброй (или гиперкомплексной системой) над полем P, если

а) для каждого элемента a поля P и каждой величины α системы S определен элемент этой системы, называемый произведением a на α и обозначаемый через $a\alpha$;

б) для каждых двух величин α, β системы однозначно определена некоторая величина той же системы, называемая суммой первых двух величин и обозначаемая через $\alpha+\beta$;

в) для каждых двух величин системы α, β однозначно определена величина той же системы, называемая произведением первых величин и обозначаемая через $\alpha\beta$;

и если указанные три действия обладают следующими свойствами[1]:

1′) $\alpha+\beta=\beta+\alpha$,

2′) $(\alpha+\beta)+\gamma=\alpha+(\beta+\gamma)$,

3′) в системе S существует нулевая величина 0 со свойством

$$\alpha+0=\alpha,$$

4′) $a(\alpha+\beta)=a\alpha+a\beta$,

5′) $(a+b)\alpha=a\alpha+b\alpha$,

6′) $(ab)\alpha=a(b\alpha)$,

7′) $0\alpha=0$, $1\cdot\alpha=\alpha$, где 1 — единичный элемент поля P,

8′) среди величин системы S существуют такие величины $\alpha_1, \alpha_2, \ldots, \alpha_n$, через которые каждая величина системы может быть однозначно представлена в виде $a_1\alpha_1+a_2\alpha_2+\ldots+a_n\alpha_n$,

9′) $(a\alpha)\beta=\alpha(a\beta)=a(\alpha\beta)$,

10′) $\alpha(\beta+\gamma)=\alpha\beta+\alpha\gamma$, $(\beta+\gamma)\alpha=\beta\alpha+\gamma\alpha$.

В этом определении роль, которую до сих пор играли действительные числа, играют элементы произвольного поля P. Из условия 8′ видно, что каждая гиперкомплексная величина определяется системой n элементов $a_1, a_2, \ldots, a_n$ поля P, и, следовательно, в зависимости от выбора поля P может определяться n комплексными числами, n рациональными числами, n действительными числами и т. п.

Первые восемь требований означают, что S образует линейное конечномерное пространство (см. главу XVI, § 2) над полем P, называемым основным полем алгебры.

Требования 9′ и 10′ можно объединить в виде равенств

$$(a\beta+b\gamma)\alpha=a(\beta\alpha)+b(\gamma\alpha),$$

$$\alpha(a\beta+b\gamma)=a(\alpha\beta)+b(\alpha\gamma),$$

из которых следует, что действие умножения есть действие линейное относительно каждого из сомножителей.

Из двух терминов «гиперкомплексная система» и «алгебра» в последние годы отдается предпочтение второму, так как элементы столь общих «гиперкомплексных систем» по своим свойствам могут настолько

[1] Буквами греческого алфавита обозначены произвольные величины системы S, а буквами латинского алфавита — элементы поля P.

сильно отличаться от обычных чисел, что называть их «гиперкомплексными числами» нецелесообразно. Термины «гиперкомплексные системы», «гиперкомплексные числа» применяются теперь лишь к простейшим алгебрам, например к системе обыкновенных кватернионов.

Из требований 1′—10′ видно, что в алгебрах не предполагается коммутативности и ассоциативности умножения, не предполагается существования единичного элемента и выполнимости «деления».

В каждой алгебре S существует базис, т. е. такая система элементов $\alpha_1, \alpha_2, \ldots, \alpha_n$, через которую все элементы алгебры однозначно представляются в виде линейных комбинаций $a_1\alpha_1 + a_2\alpha_2 + \ldots + a_n\alpha_n$ с коэффициентами из основного поля P. Каждая алгебра может иметь бесчисленное множество базисов, но число элементов каждого базиса одно и то же и называется рангом алгебры.

Система комплексных чисел, рассматриваемая как алгебра над полем действительных чисел, имеет своим базисом числа 1 и i. Но пары чисел 2 и $3i$, 1 и $a + bi$ (a, b — действительные, $b \neq 0$) также могут служить базисами.

Пусть $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n$ — базис какой-нибудь алгебры над некоторым полем P. Согласно определению, всякий элемент алгебры однозначно записывается в форме

$$\alpha = a_1\varepsilon_1 + a_2\varepsilon_2 + \ldots + a_n\varepsilon_n.$$

Если $\beta = b_1\varepsilon_1 + \ldots + b_n\varepsilon_n$ — какой-либо другой ее элемент, то, в силу свойств 1′—6′, имеем

$$\alpha + \beta = (a_1 + b_1)\,\varepsilon_1 + (a_2 + b_2)\,\varepsilon_2 + \ldots + (a_n + b_n)\,\varepsilon_n.$$

Аналогично для любого a из поля P

$$a\alpha = aa_1\varepsilon_1 + aa_2\varepsilon_2 + \ldots + aa_n\varepsilon_n.$$

Следовательно, действия сложения величин алгебры и их умножения на элементы поля P производятся вполне однозначно по приведенным формулам. Действие перемножения величин алгебры должно каждый раз задаваться особо, причем нет нужды знать, как перемножаются произвольные величины алгебры, достаточно знать закон перемножения лишь базисных величин ε_i. Действительно, в силу свойств 9′ и 10′

$$(a_1\varepsilon_1 + a_2\varepsilon_2 + \ldots + a_n\varepsilon_n)(b_1\varepsilon_1 + b_2\varepsilon_2 + \ldots + b_n\varepsilon_n) = \sum a_i b_j \cdot \varepsilon_i\varepsilon_j.$$

Каждое из произведений $\varepsilon_i\varepsilon_j$ есть некоторая величина алгебры и поэтому может быть выражена через базисные элементы

$$\varepsilon_i\varepsilon_j = c_{ij1}\varepsilon_1 + c_{ij2}\varepsilon_2 + \ldots + c_{ijn}\varepsilon_n.$$

Здесь c_{ijk} означают элементы основного поля P, над которым строится алгебра. Первый индекс означает номер первого множителя, второй —

второго множителя, а третий указывает номер того элемента, коэффициентом при котором является c_{ijk}. Коэффициенты c_{ijk} называются структурными константами алгебры, так как знание их вполне определяет все действия над величинами алгебры.

Легко подсчитать число структурных констант алгебры ранга n. Каждая константа имеет три номера i, j, k. Поэтому число структурных констант алгебры ранга n равно числу троек, образованных натуральными числами 1, 2, ..., n, т. е. равно n^3. Например, система комплексных чисел над полем действительных чисел имеет базис состоящий из чисел 1, i. В силу равенств

$$1 \cdot 1 = 1 \cdot 1 + 0 \cdot i, \qquad i \cdot 1 = 0 \cdot 1 + 1 \cdot i,$$
$$1 \cdot i = 0 \cdot 1 + 1 \cdot i, \qquad i \cdot i = -1 \cdot 1 + 0 \cdot i,$$

структурные константы будут равны соответственно

$$c_{111} = 1 \quad c_{112} = 0, \quad c_{211} = 0, \quad c_{212} = 1,$$
$$c_{121} = 0, \quad c_{122} = 1, \quad c_{221} = -1, \quad c_{222} = 0.$$

Обратно, пусть дано n^3 элементов какого-нибудь поля P, занумерованных тройками натуральных чисел c_{ijk} (i, j, $k = 1, 2, \ldots, n$). Тогда их можно принять в качестве структурных констант алгебры над полем P, принимая равенства $\varepsilon_i \varepsilon_j = \sum_{k=1}^{n} c_{ijk} \varepsilon_k$ как определение правила умножения в алгебре.

Выше мы видели, что каждая алгебра, вообще говоря, имеет бесчисленное множество различных базисов. Структурные константы зависят от выбора базиса, и поэтому одна и та же алгебра определяется различными системами структурных констант.

Какие же алгебры следует считать различными и какие одинаковыми? В теории алгебр принято считать две алгебры над одним и тем же полем P одинаковыми, если они изоморфны, т. е. если величины одной алгебры можно так взаимно однозначно сопоставить с величинами другой, что сумма и произведение двух любых величин первой алгебры будут сопоставлены соответственно с суммой и произведением соответствующих величин второй алгебры, а произведению какого-либо элемента поля P на величину из первой алгебры будет отвечать произведение того же элемента поля P на соответственный элемент второй алгебры.

Это определение одинаковости алгебр показывает, что в теории алгебр изучают лишь те свойства величин и систем величин алгебр, которые находят свое выражение в виде некоторых свойств трех основных операций. Короче говоря, теория алгебр изучает свойства операций, производимых над величинами алгебр, и не интересуется природой величин, составляющих алгебры.

Легко доказать, что если две алгебры изоморфны, то величинам, составляющим базис одной алгебры, отвечают величины, образующие базис другой, причем структурные константы, вычисленные в соответствующих базисах, являются соответственно равными. Обратно, если две алгебры над одним и тем же полем имеют в подходящих базисах соответственно равные структурные константы, то такие алгебры изоморфны.

Среди алгебр весьма важную роль играли и до сих пор играют ассоциативные алгебры, т. е. алгебры, действие умножения в которых удовлетворяет ассоциативному закону $\alpha(\beta\gamma) = (\alpha\beta)\gamma$. Изложению свойств таких алгебр и посвящен настоящий параграф. Среди неассоциативных наиболее интересными являются алгебры Ли, для которых предполагается выполнение следующих свойств умножения:

$$\alpha\beta = -\beta\alpha, \quad \alpha(\beta\gamma) + \beta(\gamma\alpha) + \gamma(\alpha\beta) = 0.$$

Они представляют интерес ввиду тесной связи, существующей между алгебрами Ли и группами Ли, о которых шла речь в § 7.

Алгебра матриц. Выше указывалось, что в продолжение первого периода развития теории гиперкомплексных систем главное внимание обращалось на исследование отдельных систем, по тем или иным причинам вызывавших особый интерес исследователей. Некоторые из этих систем были уже нами разобраны. Примерно в середине прошлого века начались исследования алгебры матриц, играющей ныне основную роль в общей теории алгебр. Мы здесь напомним кратко определения действий с матрицами (см. главу XVI, § 1).

Матрицей над полем P называется совокупность элементов этого поля, расположенных в виде прямоугольной таблицы. Две матрицы называются равными, если равны их элементы, стоящие на соответственных местах. Здесь мы будем рассматривать только квадратные матрицы, число строк которых равно числу их столбцов. Число строк квадратной матрицы или равное ему числу столбцов называется порядком матрицы.

Чтобы сложить две матрицы одинакового порядка, складывают их соответственные элементы. Умножение числа на матрицу по определению означает умножение на это число всех элементов матрицы. Действие умножения матрицы на матрицу определяется более сложно: произведением двух матриц порядка n называется матрица того же порядка, у которой элемент, стоящий в i-й строке и j-м столбце, равен сумме произведений элементов i-й строки первой матрицы на соответственные элементы j-го столбца второй. Например:

$$\begin{pmatrix} a & b \\ a_1 & b_1 \end{pmatrix} \begin{pmatrix} x & y \\ x_1 & y_1 \end{pmatrix} = \begin{pmatrix} ax + bx_1 & ay + by_1 \\ a_1x + b_1x_1 & a_1y + b_1y_1 \end{pmatrix}.$$

Причины, по которым определение перемножения матриц выбрано именно так, были изложены в главе XVI.

В силу указанных определений матрицы n-го порядка с элементами из какого-либо поля P образуют систему величин, которые можно складывать, умножать на элементы поля P и перемножать между собой. Несложные вычисления показывают, что свойства 1′—10′, определяющие алгебру, здесь выполняются. Кроме того, легко доказывается, что умножение матриц подчиняется ассоциативному закону. Поэтому система всех матриц данного порядка n с элементами из заданного поля P образует ассоциативную алгебру над этим полем.

Очевидно равенство

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} = a\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} + b\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} + c\begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} + d\begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$$

показывает, что четыре матрицы, стоящие в правой части, образуют базис алгебры матриц 2-го порядка. Вообще, обозначая через ε_{ij} матрицу, у которой в i-й строке и j-м столбце стоит 1, а остальные места заняты нулями, будем иметь равенство

$$\begin{pmatrix} a_{11} & \dots & a_{1n} \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \cdot & & \cdot \\ a_{n1} & \dots & a_{nn} \end{pmatrix} = \sum_{i,\,j} a_{ij}\varepsilon_{ij},$$

показывающее, что матрицы ε_{ij} образуют базис алгебры матриц n-го порядка. Так как число матриц ε_{ij} равно n^2, то ранг алгебры матриц также равен n^2. Таблица умножения базисных матриц ε_{ij} имеет вид

$$\varepsilon_{ij} \cdot \varepsilon_{jl} = \varepsilon_{il}, \quad \varepsilon_{ij} \cdot \varepsilon_{kl} = 0, \quad j \neq k, \quad i, j, k, l = 1, 2, \dots, n.$$

Алгебра матриц содержит единицу, роль которой играет единичная матрица.

Представления ассоциативных алгебр. Пусть каждой величине некоторой алгебры A над полем P отнесена определенная величина какой-либо алгебры B над тем же полем P. Если при этом сумме и произведению каждых двух элементов алгебры A отвечают сумма и произведение соответствующих элементов алгебры B и произведению каждого элемента поля P на произвольный элемент алгебры A отвечает произведение того же элемента поля P на соответствующий элемент алгебры B, то говорят, что алгебра A отображена гомоморфно в алгебру B. Гомоморфное отображение ассоциативной алгебры в алгебру матриц порядка n называется представлением алгебры A степени n. Если различным элементам алгебры A отвечают различные матрицы, то представление называется точным или изоморфным. Когда алгебра A представлена изоморфно матрицами, можно считать, что действия над величинами алгебры сводятся к действиям над соответствующими матрицами. Поэтому задача нахождения представлений

алгебр имеет значительный интерес. Мы рассмотрим здесь только самые простые способы нахождения представлений алгебр, которые, однако, играют важную роль в общей теории.

Выберем в заданной ассоциативной алгебре A какой-либо базис $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n$ и пусть α — произвольная величина из A. Произведения $\varepsilon_1\alpha, \varepsilon_2\alpha, \ldots, \varepsilon_n\alpha$ являются снова величинами из A и потому должны выражаться линейно через $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n$. Пусть

$$\begin{aligned}
\varepsilon_1\alpha &= a_{11}\varepsilon_1 + a_{12}\varepsilon_2 + \ldots + a_{1n}\varepsilon_n,\\
\varepsilon_2\alpha &= a_{21}\varepsilon_1 + a_{22}\varepsilon_2 + \ldots + a_{2n}\varepsilon_n,\\
&\ldots\ldots\ldots\ldots\ldots\ldots\\
\varepsilon_n\alpha &= a_{n1}\varepsilon_1 + a_{n2}\varepsilon_2 + \ldots + a_{nn}\varepsilon_n.
\end{aligned}$$

Как видим, при фиксированном базисе с каждым элементом α может быть сопоставлена определенная матрица $\|a_{ij}\|$. Весьма простой подсчет показывает, что это сопоставление является представлением алгебры A. Это представление часто называют регулярным представлением алгебры A. Степень его очевидно равна рангу алгебры.

Комплексные числа можно рассматривать как алгебру ранга 2 над полем действительных чисел с базисом 1, i. Равенства

$$\begin{aligned}
1\cdot(a+bi) &= a\cdot 1 + b\cdot i,\\
i\cdot(a+bi) &= -b\cdot 1 + a\cdot i
\end{aligned}$$

показывают, что в соответствующем регулярном представлении комплексному числу $a+bi$ отвечает матрица $\begin{pmatrix} a & b\\ -b & a\end{pmatrix}$. Аналогичное представление кватернионов имеет вид

$$a+bi+cj+dk \to \begin{pmatrix} a & b & c & d\\ -b & a & -d & c\\ -c & d & a & -b\\ -d & -c & b & a\end{pmatrix}.$$

Указанные представления комплексных чисел и кватернионов являются точными (т. е. изоморфными самой алгебре). Примеры показывают, однако, что регулярное представление не всегда точное. Но если алгебра содержит единицу, то ее регулярное представление заведомо точное.

Легко показать, что всякую ассоциативную алгебру можно включить в алгебру с единицей. Регулярное представление объемлющей алгебры будет точным; следовательно, это представление будет точным и для данной алгебры. Таким образом, всякая ассоциативная алгебра обладает точным представлением матрицами.

Указанный способ нахождения представлений недостаточен для построения всех представлений алгебры. Более тонкий способ связан

с понятием идеала алгебры, которое вообще играет большую роль в современной математике.

Система I элементов некоторой алгебры называется правым *идеалом*, если она есть линейное подпространство алгебры и произведение любого элемента из I на любой элемент алгебры снова содержится в I. Аналогично (с переменой порядка множителей) определяется левый идеал. Идеал, являющийся одновременно левым и правым, называется двусторонним. Ясно, что нулевой элемент алгебры сам по себе уже образует двусторонний идеал — так называемый нулевой идеал алгебры. Так же и вся алгебра может рассматриваться как свой собственный двусторонний идеал. Однако, кроме этих двух тривиальных идеалов, алгебра может содержать и другие идеалы, существование которых обычно связано с интересными свойствами алгебры.

Пусть ассоциативная алгебра A содержит правый идеал I. Выберем в этом идеале базис $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_m$. Так как I составляет в общем случае лишь часть A, то базис I будет иметь, как правило, меньше элементов, чем базис A. Пусть α — произвольный элемент из A. Так как I правый идеал и $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_m$ содержатся в I, то произведения $\varepsilon_1\alpha, \ldots, \varepsilon_m\alpha$ также содержатся в I и, значит, выражаются линейно через базис $\varepsilon_1, \ldots, \varepsilon_m$, т. е.

$$\begin{aligned} \varepsilon_1\alpha &= a_{11}\varepsilon_1 + \ldots + a_{1m}\varepsilon_m, \\ &\cdots\cdots\cdots\cdots \\ \varepsilon_m\alpha &= a_{m1}\varepsilon_1 + \ldots + a_{mm}\varepsilon_m. \end{aligned}$$

Сопоставляя с элементом α матрицу $\|a_{ij}\|$, получим, как и ранее представление алгебры A. Степень этого представления равна числу элементов базиса идеала I и, следовательно, в общем случае будет меньшей, чем степень регулярного представления. Очевидно, степень представления, полученного при помощи идеала, будет наименьшей, если сам идеал будет минимальным. Отсюда можно понять особую роль минимальных идеалов в теории алгебр.

Строение алгебр. Согласно сказанному каждая ассоциативная алгебра A может быть изоморфно представлена матрицами некоторого порядка. Совокупность матриц, отвечающих в этом представлении величинам алгебры A, будет сама алгеброй, но являющейся лишь частью алгебры всех матриц данного порядка. Если некоторая часть величин алгебры сама является алгеброй, то она называется подалгеброй данной алгебры. Можно сказать, следовательно, что всякая ассоциативная алгебра изоморфна некоторой подалгебре матриц.

Хотя этот результат представляет принципиальный интерес, так как он сводит вопрос о нахождении всех алгебр к нахождению всевозможных подалгебр матричных алгебр, он не дает прямого ответа на вопрос о строении алгебр. Впервые общий ответ на этот вопрос был дан в конце прошлого века в работах профессора Юрьевского

(Тартуского) университета Ф. Э. Молина (1861—1941), с 1900 г. преподававшего в Томском политехническом институте.

Алгебру называют *простой*, если она не содержит никаких двусторонних идеалов, отличных от нулевого и всей алгебры. Ф. Э. Молиным было показано, что всякая простая ассоциативная алгебра ранга 2 или большего над полем комплексных чисел изоморфна алгебре всех матриц подходящего порядка над этим полем.

Продолжая основополагающие исследования Молина, Веддербарн получил в начале XX в. ряд результатов, вскрывающих весьма полно строение алгебр над произвольным полем.

Какая-либо система элементов алгебры A (в частности сама алгебра A, ее некоторый идеал или подалгебра) называется *нильпотентной*, если существует такое натуральное число s, что произведение любых s элементов системы равно нулю. Всякая ассоциативная алгебра обладает единственным максимальным двусторонним нильпотентным идеалом, называемым радикалом алгебры. Алгебра, радикал которой равен нулю, называется *полупростой*. Можно показать, что всякая полупростая алгебра распадается в особого рода сумму простых алгебр, благодаря чему изучение полупростых алгебр целиком сводится к изучению простых. Наконец, алгебра A называется *алгеброй с делением*, если в A каждое уравнение вида $ax = b\,(a \neq 0)$ имеет решение.

Строение простых алгебр над полем комплексных чисел полностью описывается упомянутой теоремой Молина. Если же основное поле P произвольно, то имеет место более общая теорема Веддербарна: всякая простая алгебра ранга 2 или большего над полем P изоморфна алгебре всех матриц подходящего порядка с элементами из некоторой алгебры с делением над тем же полем P. Таким образом, теорема Веддербарна сводит вопрос о нахождении простых алгебр над заданным полем P к нахождению алгебр с делением над полем P. Над полем комплексных чисел есть только одна алгебра с делением — само поле комплексных чисел. По теореме Веддербарна отсюда следует, что все простые алгебры над полем комплексных чисел изоморфны алгебрам матриц над этим полем, т. е. следует теорема Молина.

Над полем действительных чисел существуют лишь три ассоциативные алгебры с делением: само поле действительных чисел, поле комплексных чисел и алгебра кватернионов. Доказательство этого утверждения не очень просто, и мы не будем на нем останавливаться. В силу теоремы Веддербарна отсюда вытекает, что каждая простая алгебра над полем действительных чисел изоморфна алгебре матриц подходящего порядка либо над полем действительных чисел, либо над полем комплексных чисел, либо над алгеброй кватернионов.

Из этих примеров видно, как раскрывается строение полупростых алгебр теоремами Молина и Веддербарна. Что касается алгебр с радикалом, то для них большое значение имеет так называемая основная

теорема Веддербарна, согласно которой при некоторых ограничениях, накладываемых на основное поле, в каждой алгебре A с радикалом R существует полупростая подалгебра L, такая, что каждый элемент заданной алгебры может быть однозначно представлен в виде суммы $\lambda+\rho$ ($\lambda\in L$, $\rho\in R$), причем подалгебра L определяется в некотором смысле однозначно внутри алгебры A.

Только что сформулированные основные теоремы дают стройное представление о возможных типах ассоциативных алгебр и сводят вопрос об их строении в основном к аналогичному вопросу о строении нильпотентных алгебр. Теория последних пока еще находится в процессе становления.

§ 13. АЛГЕБРЫ ЛИ

В § 12 говорилось, что, кроме теории ассоциативных алгебр, в настоящее время весьма детально разработана теория алгебр Ли, умножение в которых подчинено требованиям

$$\alpha\beta=-\beta\alpha, \qquad \alpha(\beta\gamma)+\beta(\gamma\alpha)+\gamma(\alpha\beta)=0.$$

Важность этих алгебр объясняется тем, что они тесно связаны с группами Ли (см. § 7), т. е. с важнейшим классом непрерывных групп. Как мы видели выше, группы Ли играют значительную роль в современной геометрии. В соответствии с происхождением теории групп и алгебр Ли наибольший интерес представляют алгебры Ли над полями всех действительных и всех комплексных чисел

Одним из простых примеров алгебры Ли является следующий. Рассмотрим множество всех квадратных матриц данного порядка n. Введем для них действие коммутирования, понимая под ним составление по данным матрицам A и B их так называемого коммутатора $AB-BA$, обозначаемого через $[A, B]$.

Нетрудно проверить, что

$$[A, B]=-[B, A],$$
$$[A, [B, C]]+[B, [C, A]]+[C, [A, B]]=0.$$

Следовательно, множество всех квадратных матриц данного порядка образует алгебру Ли относительно операции коммутирования. Ясно, что всякая подалгебра алгебры Ли, образованной матрицами, т. е. всякое множество матриц, замкнутое относительно действий сложения, умножения на числа основного поля и коммутирования, в свою очередь является алгеброй Ли.

Вопрос о том, для всякой ли абстрактно заданной алгебры Ли существует изоморфная ей матричная алгебра, долгое время оставался открытым. Он был решен положительно только в 1935 г. И. Д. Адо, учеником известного алгебраиста Н. Г. Чеботарева.

Коснемся теперь в общих чертах, не входя в подробности и не давая строгих формулировок, соответствия между группами Ли и алгебрами Ли, ограничившись случаем, когда группа Ли и алгебра Ли представлены матрицами.

Пусть L — некоторая матричная алгебра Ли. Сопоставим каждой матрице A, принадлежащей L, матрицу $U = e^A = E + \frac{A}{1!} + \frac{A^2}{2!} + \ldots$

Тогда совокупность всех полученных таким образом матриц образует группу Ли относительно обычного матричного умножения. Обратно, для каждой группы Ли найдется единственная (с точностью до изоморфизма) алгебра Ли, такая, что соответствующая ей группа будет изоморфна данной.

Для простоты мы привели не точную, а упрощенную формулировку теоремы о соответствии между группами и алгебрами Ли. В действительности соотношение $U = e^A$ существует лишь для U, достаточно близкой к единичной матрице, и A, достаточно близкой к нулевой матрице. Строгая формулировка требует введения довольно сложных понятий локальной группы и локального изоморфизма.

Таким образом, переход от алгебры Ли к соответствующей группе осуществляется посредством действия, аналогичного потенцированию, а обратный переход — от группы к алгебре — посредством действия, аналогичного логарифмированию.

Если L совпадает с алгеброй всех матриц порядка n, то соответствующая группа Ли будет группой всех неособенных матриц, так как любая близкая к единичной матрица U может быть представлена в виде $U = e^A$.

Матрица $A = \| a_{ij} \|$ называется кососимметрической, если ее элементы удовлетворяют соотношению $a_{ji} = -a_{ij}$. Кососимметрические матрицы образуют алгебру Ли, так как если A и B кососимметрические, то матрицы $AB - BA = [A, B]$ и $\alpha A + \beta B$ будут также кососимметрическими.

Легко проверить, что для каждой кососимметрической матрицы A выражение e^A будет ортогональной матрицей, причем каждая ортогональная матрица, близкая к единичной, может быть представлена в указанной экспоненциальной форме. Следовательно, алгеброй Ли группы ортогональных матриц является алгебра кососимметрических матриц.

Из аналитической геометрии известно, что каждое вращение пространства около начала координат задается ортогональной матрицей, причем произведению вращений отвечает произведение соответствующих матриц. Иными словами, группа вращений пространства около неподвижной точки изоморфна группе ортогональных матриц 3-го порядка. Отсюда мы заключаем, что алгебра Ли для группы вращений про-

странства есть алгебра всех кососимметрических матриц 3-го порядка, т. е. алгебра Ли матриц вида

$$A=\begin{pmatrix} 0 & -a & -b \\ a & 0 & -c \\ b & c & 0 \end{pmatrix}.$$

Так как каждая из этих матриц вполне характеризуется тремя числами a, b, c, то ее можно условно изобразить вектором $\boldsymbol{a}$, имеющим по осям координат проекции a, b, c. При этом линейной комбинации $\alpha A_1+\beta A_2$ матриц A_1 и A_2 указанного вида будет, очевидно, отвечать линейная комбинация соответствующих векторов $\alpha\boldsymbol{a}_1+\beta\boldsymbol{a}_2$, а коммутатору матриц

$$[A_1,\ A_2]=A_1A_2-A_2A_1=$$
$$=\begin{pmatrix} 0 & -a_1 & -b_1 \\ a_1 & 0 & -c_1 \\ b_1 & c_1 & 0 \end{pmatrix}\begin{pmatrix} 0 & -a_2 & -b_2 \\ a_2 & 0 & -c_2 \\ b_2 & c_2 & 0 \end{pmatrix}-\begin{pmatrix} 0 & -a_2 & -b_2 \\ a_2 & 0 & -c_2 \\ b_2 & c_2 & 0 \end{pmatrix}\begin{pmatrix} 0 & -a_1 & -b_1 \\ a_1 & 0 & -c_1 \\ b_1 & c_1 & 0 \end{pmatrix}=$$
$$=\begin{pmatrix} 0 & b_2c_1-b_1c_2 & a_1c_2-a_2c_1 \\ b_1c_2-b_2c_1 & 0 & a_2b_1-a_1b_2 \\ a_2c_1-a_1c_2 & a_1b_2-a_2b_1 & 0 \end{pmatrix}.$$

будет отвечать вектор с проекциями $b_1c_2-b_2c_1$, $a_2c_1-a_1c_2$, $a_1b_2-a_2b_1$, т. е. векторное произведение векторов $\boldsymbol{a}_1$ и $\boldsymbol{a}_2$. Мы пришли к замечательному результату, что совокупность обычных векторов относительно действий сложения, умножения на скаляр и векторного умножения образует алгебру Ли, отвечающую группе вращений пространства около неподвижной точки. Это еще раз показывает, насколько тесно геометрические понятия связаны с группой вращения пространства, т. е., другими словами, с законами движений твердых тел.

Для алгебр Ли в конце прошлого и начале нынешнего столетия был получен ряд результатов, аналогичных основным результатам об ассоциативных алгебрах, хотя доказательства и формулировки здесь оказываются более сложными. Так, в результате усилий Ли, Киллинга и Картана к началу XX в. для алгебр Ли удалось установить понятия радикала, полупростоты и найти все простые алгебры Ли над полями действительных и комплексных чисел. К началу 30-х годов Картаном и Вейлем была в основном построена теория представлений алгебр Ли матрицами, оказавшаяся замечательным инструментом для решения многих задач. В последние 15 лет разработкой теории алгебр Ли занимался ряд советских математиков, получивших в этой области немало замечательных результатов. В частности, ими была существенно продвинута теория представлений алгебр Ли и окончательно решены вопросы о полупростых подалгебрах алгебр Ли, о построении алгебр с заданным радикалом и т. п.

§ 14. КОЛЬЦА

В § 11 (стр. 305) было дано общее определение поля как произвольного множества элементов, на котором определены действия сложения и умножения, удовлетворяющие приводившимся там требованиям 1—10. Опустив в этом определении требование 10 о существовании частного и требования 7, 8 коммутативности и ассоциативности умножения, получим определение понятия кольца — одного из важнейших понятий современной алгебры.

Всякое поле, а также всякая алгебра, рассматриваемая только относительно операций сложения и умножения, является кольцом. Еще более простым примером кольца служит совокупность целых рациональных чисел с обычными операциями сложения и умножения. Относительно этих операций кольцами будут также совокупности чисел вида $a+bi$, $a+b\sqrt{2}$, $a+b\sqrt[3]{2}+c\sqrt[3]{4}$ и т. п., где a, b, c — целые рациональные числа. Элементами этих колец являются числа, благодаря чему и сами кольца называются числовыми. Некоторые важные свойства и приложения этих колец были рассмотрены в главах IV (том 1) и X (том 2).

Однако существуют важные классы и нечисловых колец. Так, например, относительно обычных операций сложения и умножения кольцами являются совокупность многочленов от данных переменных $x_1, x_2, \ldots, x_n$ с коэффициентами из какого-либо фиксированного кольца или поля, совокупность всех непрерывных функций, определенных на некоторой области, совокупность линейных преобразований линейного или гильбертова пространств.

Арифметические свойства числовых колец являются предметом изучения глубокой теории алгебраических чисел, пограничной между собственно алгеброй и собственно теорией чисел. Исследованием свойств колец многочленов занимается так называемая теория полиномиальных идеалов, тесно связанная с высшими отделами аналитической геометрии. Наконец, кольца функций и преобразований играют основную роль в функциональном анализе (см. главу XIX).

На базе этих и некоторых других конкретных теорий в текущем столетии стали быстро развиваться общая теория колец и теория топологических колец.

За ограниченностью места далее будут указаны лишь отдельные результаты, относящиеся только к введению в теорию колец.

Идеалы. Подмножество I элементов некоторого (не обязательно ассоциативного) кольца K называется его *идеалом*, если разность любых двух элементов из I снова содержится в I и если произведения ax, xa произвольного элемента a из I на произвольный элемент x кольца K содержатся в I.

Каждый идеал является такой частью кольца, которая сама является кольцом относительно действующих в заданном кольце операций сложения и умножения. Такие части называются подкольцами данного кольца и, значит, каждый идеал является в то же время подкольцом. Обратное, как правило, несправедливо.

Пересечение любой системы идеалов кольца снова является идеалом этого кольца, в частности идеалом кольца будет пересечение всех идеалов, содержащих какой-нибудь фиксированный элемент a кольца. Этот идеал называется главным идеалом, порожденным элементом a, и обозначается через (a).

Таким же образом определяется понятие идеала, порожденного двумя или несколькими элементами. Легко показать, что если ассоциативное коммутативное кольцо имеет единичный элемент, то идеалом, порожденным элементами $a_1, \ldots, a_n$, будет просто совокупность всех элементов кольца, допускающих запись в виде суммы $x_1a_1 + \ldots + x_na_n$, где $x_1, \ldots, x_n$ — произвольные элементы кольца. В частности, главный идеал (a) в коммутативном ассоциативном кольце с единицей есть просто совокупность всех элементов, кратных a, т. е. имеющих вид xa.

В кольце всех целых рациональных чисел каждый идеал является главным. Тем же свойством обладают кольцо многочленов от одного переменного с коэффициентами из некоторого поля, кольцо комплексных чисел вида $a + bi$, где a, b — целые рациональные, и ряд других колец. Однако уже совокупность всех многочленов от двух переменных x, y без свободного члена не будет главным идеалом в кольце всех многочленов от x, y с рациональными коэффициентами.

Аналогично тому, как это было с нормальными делителями в теории групп, для каждого идеала I кольца K можно построить фактор-кольцо K/I. Делается это следующим образом. Элементы a, b кольца K называются сравнимыми по идеалу I, символически $a \equiv b\ (I)$, если их разность $a - b$ содержится в I. Легко устанавливается, что отношение сравнимости симметрично, рефлексивно и транзитивно (см. главу XV) и, следовательно, все элементы K разбиваются на классы сравнимых между собой по идеалу I. Рассматривая теперь эти классы как элементы нового множества, вводят для них понятие суммы и произведения, называя «суммой» двух классов тот класс, который содержит сумму каких-либо двух элементов, входящих соответственно в эти классы, а произведением — класс, содержащий произведение указанных представителей. Из определения идеалов следует, что так определенные сумма и произведение на самом деле не зависят от выбора представителей и что в результате совокупность классов становится кольцом.

Роль фактор-колец в теории колец совершенно аналогична роли фактор-групп в теории групп. В частности, построение фактор-колец

от известных колец представляет собою удобный способ образования колец с самыми различными свойствами. Более того, легко доказывается, например, что произвольное коммутативное кольцо K изоморфно фактор-кольцу кольца многочленов с целыми рациональными коэффициентами от достаточного числа переменных.

Арифметические свойства колец. В числовых кольцах и в полях произведение нескольких элементов может равняться нулю, только если хотя бы один из сомножителей равен нулю. В произвольных кольцах это может оказаться неверным, например произведение двух ненулевых матриц может быть равно нулю. Если в некотором кольце $ab = 0$, причем $a \neq 0$, $b \neq 0$, то a и b называются *делителями нуля*. Если таких элементов в кольце нет, то кольцо называется *кольцом без делителей нуля*.

При исследовании законов делимости в кольцах обычно предполагается, что кольцо коммутативно и не имеет делителей нуля. Такие кольца принято называть областями целостности. Упомянутые выше числовые и полиномиальные кольца являются областями целостности.

Пусть K — некоторая область целостности. Говорят, что элемент a делится в области K на элемент b, если $a = bq$, $q \in K$. Отсюда непосредственно следует, что сумма элементов, делящихся на b, делится на b и что произведение нескольких элементов из K заведомо делится на b, если один из сомножителей делится на b. При введении понятия простого элемента, аналогичного понятию простого числа, в теории колец возникает усложнение, уже упоминавшееся в главе X (том 2). Именно, сначала приходится ввести понятие ассоциированных элементов кольца, называя элементы a, b ассоциированными, если a делится на b, а b делится на a. Полагая $a = bq_1$, $b = aq_2$, имеем $ab = ab \cdot q_1 q_2$, т. е. $q_1 q_2 = e$, где e — единица области K. Частные ассоциированных элементов называются поэтому делителями единицы. Всякий элемент области делится на любой делитель единицы. В кольце целых рациональных чисел делителями единицы являются ± 1, в кольце чисел вида $a + bi$, где a, b — целые, делителями единицы будут числа ± 1, $\pm i$.

Каждый элемент области целостности K обладает разложениями вида $a = a\varepsilon \cdot \varepsilon^{-1}$, где ε — какой-либо делитель единицы. Эти разложения называются тривиальными. Если никаких других разложений у a нет, то a называется простым или неразложимым элементом K. В связи с важным значением теоремы об однозначном разложении целых чисел на простые множители представляет интерес нахождение таких классов колец, в том числе и некоммутативных, в которых остается справедливой аналогичная теорема. Например эта теорема имеет место в кольцах главных идеалов, т. е. областях целостности, в которых все идеалы главные.

В связи с вопросом об однозначности разложения на простые сомножители возникло и само понятие об идеалах. Приблизительно

в середине прошлого века немецкий математик Куммер, пытаясь доказать знаменитое предположение Ферма о том, что уравнение $x^n+y^n=z^n$ не имеет ненулевых целочисленных решений при $n\geqslant 3$, пришел к мысли рассматривать числа вида $a_0+a_1\zeta+\ldots+a_n\zeta^{n-1}$, где $\zeta=\cos\frac{2\pi}{n}+i\sin\frac{2\pi}{n}$ — решение уравнения $x^n=1$, а $a_0,\ldots,a_n$ — обычные целые числа. Числа указанного вида образуют область целостности, и Куммер сначала принял в качестве очевидного предположение, что в этой области имеет место теорема об однозначности разложения на простые множители. При этом им было построено и доказательство предположения Ферма. Однако при проверке обнаружилось, что упомянутое допущение об однозначности разложения неверно. Желая сохранить однозначность разложения на простые сомножители, Куммер оказался вынужденным рассматривать разложения чисел области на сомножители, не входящие в самую область. Эти числа он назвал идеальными. Впоследствии при построении общей теории вместо идеальных чисел стали вводить совокупности элементов области, делящиеся на то или иное идеальное число, которые и получили название идеалов.

Открытие неоднозначности разложения на простые сомножители в числовых кольцах представляется одним из интереснейших фактов, найденных в прошлом столетии, приведшим к созданию обширной теории алгебраических чисел.

Одно из изящных применений этой теории к вопросу о разложении обычных целых чисел на сумму квадратов указано в конце главы X (том 2). Значительную роль в развитии теории числовых колец сыграли работы отечественных математиков Е. И. Золотарева, Г. Ф. Вороного, И. М. Виноградова и Н. Г. Чеботарева.

Алгебраические многообразия. Другой исток теории идеалов лежит в алгебраической геометрии. Уже при первоначальном ознакомлении с теорией кривых второго порядка обычно вызывает удивление, что единой кривой гиперболой называют совокупность двух не связанных друг с другом кривых — ветвей гиперболы и что в то же время пару прямых называют распадающейся кривой второго порядка. Это различие в терминологии находит объяснение в алгебре: если уравнения кривых рассматривать в виде $f(x, y)=0$, где $f(x, y)$ — многочлен от x, y, то в первом случае левая часть этого уравнения будет неприводимым многочленом второй степени, а во втором — произведением двух сомножителей первой степени. Кривую, уравнение которой можно представить при помощи неприводимого многочлена $f(x, y)$, называют неприводимой, а в противном случае — приводимой.

При переходе к пространственным кривым дело становится сложнее. Пространственная кривая изображается системой двух уравнений $f(x, y, z)=0$, $g(x, y, z)=0$, причем многочлены f и g определяются

кривой далеко не однозначно. Что же здесь называть неприводимой кривой?

Естественный ответ дает теория идеалов. Пусть f_1, f_2, ... — некоторое множество многочленов от переменных x, y, z вообще с комплексными коэффициентами. Совокупность точек пространства (комплексного), координаты которых обращают все данные многочлены в нуль, называется алгебраическим многообразием, определяемым данными многочленами. Обозначим это многообразие через M и рассмотрим все многочлены от переменных x, y, z, обращающиеся в нуль в каждой точке M. Легко видеть, что совокупность I всех таких многочленов будет идеалом в кольце многочленов от x, y, z. Этот идеал, кроме того, будет обладать тем свойством, что если степень какого-либо многочлена содержится в I, то и сам многочлен содержится в I. Оказывается, что в то время как различные совокупности многочленов могут определять одно и то же алгебраическое многообразие, соответствие между многообразиями и идеалами с упомянутым дополнительным свойством является взаимно однозначным.

Таким образом, при изучении свойств многообразий естественно рассматривать не более или менее случайные «уравнения» их, а изучать соответствующий идеал. Если идеал I может быть представлен в виде пересечения каких-либо двух идеалов I_1, I_2, то многообразие M будет объединением многообразий M_1, M_2, отвечающих идеалам I_1, I_2. Отсюда видно, что многообразие M естественно называть неприводимым в том случае, когда соответствующий идеал I нельзя представить в виде пересечения двух других объемлющих идеалов. Распадению кривой на кривые низших порядков, разложению многообразия на неприводимые теперь будет отвечать представление соответствующего идеала в виде пересечения неразложимых. Вопрос об однозначности и возможности таких разложений является одним из первых в теории алгебраических многообразий и общей теории идеалов.

Строение некоммутативных колец. Всякая алгебра является в то же время кольцом относительно операций сложения и умножения. Поэтому значительное число основных понятий и результатов теории алгебр имеет силу и для произвольных колец. Однако перенесение более тонких результатов теории алгебр, аналогичных, в частности, теоремам Молина—Веддербарна (ср. § 11), связано с рядом трудностей, отчасти преодоленных только за последние 10—15 лет. Речь прежде всего идет о нахождении такого определения радикала кольца, чтобы кольца с нулевым радикалом имели какое-то сходство с полупростыми алгебрами, и во всяком случае, чтобы из теорем теории колец соответствующие результаты структурной теории алгебр получались в качестве частных случаев. В настоящее время в теории колец имеется уже ряд определений радикала, позволяющих при тех или иных ограничениях строить содержательную теорию строения полупростых колец

Как уже отмечалось, интерес к теории некоммутативных колец в значительной мере стимулируется весьма заметным значением теории колец операторов в функциональном анализе.

§ 15. СТРУКТУРЫ

Как читатель уже знает, множество объектов называется частично упорядоченным, если для некоторых пар его элементов определено, какой из этих объектов предшествует другому или подчинен другому, предполагая при этом, что: 1) каждый объект подчинен самому себе; 2) из подчиненности a объекту b и подчиненности b объекту a следует тождественность a с b; 3) из подчиненности a объекту b и подчиненности b объекту c следует подчиненность a объекту c. Отношение подчиненности обычно обозначают знаком $\leqslant$.

Важным примером частично упорядоченной совокупности является система всех подмножеств какого-либо множества, где отношение подчиненности означает, что одно подмножество является частью другого.

Если отношение подчиненности определено для каждой пары элементов частично упорядоченного множества, то множество называется упорядоченным (линейно). Упорядоченным, например, является множество действительных чисел, где отношение $a \leqslant b$ означает, что число a не больше b. Напротив, частично упорядоченное множество всех частей какой-либо совокупности, содержащей более одного элемента, не будет упорядоченным, так как подмножества без общих элементов уже не будут сравнимыми между собой.

Пусть элементы некоторого частично упорядоченного множества M обладают тем свойством, что каждая пара их a, b имеет единственный ближайший общий больший элемент c, т. е. такой, для которого $a \leqslant c$, $b \leqslant c$ и при любом d из M, удовлетворяющем условиям $a \leqslant d$, $b \leqslant d$, будет $c \leqslant d$. Тогда M называется *верхней полуструктурой*, а элемент c «суммой» a и b. Легко убедиться, что это «сложение» обладает следующими свойствами:

$$a+b=b+a, \quad (a+b)+c=a+(b+c), \quad a+a=a. \tag{13}$$

Весьма замечательно, что можно утверждать и обратное. Если в некотором множестве определено действие сложения, обладающее свойствами (13), то, называя элемент a подчиненным элементу b, если $a+b=b$, мы получим частично упорядоченное множество, в котором $a+b$ будет единственным ближайшим общим бóльшим для a и b.

Аналогично можно определить нижние полуструктуры, рассматривая вместо ближайших бóльших ближайшие меньшие, которые здесь называются «произведениями» данных элементов. Эта операция обладает теми же свойствами, что и «сложение», именно

$$ab=ba, \quad (ab)c=a(bc), \quad aa=a. \tag{14}$$

Частично упорядоченное множество, являющееся одновременно верхней и нижней полуструктурой, называется *структурой*. Согласно изложенному, в каждой структуре можно определить два действия, подчиненные условиям (13), (14). Однако эти действия связаны друг с другом, так как отношение $a \leqslant b$ в структуре можно записать в любой из форм $a+b=b$, $ab=a$. Иначе говоря, в структурах равенства $a+b=b$ и $ab=a$ должны быть равносильными. Оказывается, последнее условие можно записать алгебраически в виде равенств

$$a+ab=a, \quad a(a+b)=a, \tag{15}$$

благодаря чему изучение структур становится чисто алгебраической задачей об изучении систем с двумя действиями, подчиненными требованиям (13), (14), (15). Значение алгебраического подхода к изучению структур, грубо говоря, состоит в том, что особенности той или иной конкретной структуры в отдельных случаях удается выразить в виде тех или иных алгебраических соотношений между элементами, а также воспользоваться богатым аппаратом классических теорий групп и колец.

Как уже упоминалось, совокупность всех подмножеств некоторого множества является частично упорядоченным множеством. Нетрудно видеть, что оно будет структурой, причем структурной суммой будет здесь объединение, а структурным произведением — пересечение соответствующих подмножеств. Если рассматривать не все, а только некоторые подмножества, то можно получить самые разнообразные структуры. Например, структурой будет совокупность всех подгрупп, а также всех инвариантных подгрупп произвольной группы, совокупность всех подколец и совокупность всех идеалов произвольного кольца и т. п. В частности, в структурах всех инвариантных подгрупп группы и всех идеалов кольца, кроме основных тождеств (13), (14), (15), имеет место еще следующий так называемый модулярный закон

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

Теория структур с модулярным законом (дедекиндовых структур) составляет важную главу общей теории структур.

Значительное число теорем теории групп и теории колец является высказываниями о расположении подгрупп, инвариантных подгрупп и идеалов, вследствие чего эти теоремы могут быть переформулированы как теоремы о структурах подгрупп или идеалов. При некоторых ограничениях аналогичные теоремы имеют место и для общих структур. Таким путем в теорию структур были перенесены некоторые важные теоремы из теории групп, теории колец и других дисциплин. С другой стороны, использование аппарата теории структур оказалось полезным, наоборот, при нахождении свойств конкретных структур, например в теории групп и теории колец.

Теория структур возникла совсем недавно — в 20—30-х годах нашего столетия и еще не имеет таких важных применений, как, скажем, теория групп. Однако уже в настоящее время теория структур — вполне оформившаяся математическая дисциплина с большим содержанием и существенной проблематикой.

§ 16. ОБЩИЕ АЛГЕБРАИЧЕСКИЕ СИСТЕМЫ

В предыдущих параграфах была сделана попытка дать понятие о том, как применение алгебраических методов к все расширяющемуся кругу задач привело к расширению систем объектов, изучаемых алгеброй, и к обобщению понятия самих алгебраических операций. Большую роль в этом сыграло развитие аксиоматического метода, вызванное работами Н. И. Лобачевского по основаниям геометрии, а также развитие общей теории множеств.

Одним из основных итогов этого явилось постепенное выкристаллизовывание общих понятий алгебраической операции, алгебраической системы и накопление важнейших фактов, относящихся к определенным алгебраическим системам. Вместо конкретно определяемых в школьной алгебре действий, относящихся большей частью к числам, в современной алгебре исходят из общего понятия действия или операции. Именно, пусть дана некоторая система элементов S и дано правило, сопоставляющее каждой системе $a_1, a_2, \ldots, a_m$ из m элементов S, взятых в определенном порядке, вполне определенный элемент a той же системы. Тогда говорят, что на системе S задана m-членная операция и что элемент a есть результат этой операции, выполненной над элементами $a_1, a_2, \ldots, a_m$. Множество элементов, с определенными на нем одной или несколькими операциями, называется алгебраической системой. Одной из основных задач алгебры является изучение и классификация алгебраических систем. Однако в такой форме задача имеет слишком общий характер. На самом деле к настоящему времени оказались действительно важными и обладающими содержательными теориями лишь некоторые специальные алгебраические системы. Так, из систем с одним действием в глубокую математическую науку разрослась пока лишь теория групп, которой были посвящены §§ 1—10 этой главы, а из систем с двумя и большим числом действий важное значение имеют теории полей, алгебр, колец и структур. Однако число фактически рассматриваемых по тому или иному поводу алгебраических систем непрерывно растет. В то же время некоторые классические разделы алгебры, как, например, учение о гомоморфизмах, о свободных системах и свободных объединениях, о прямых объединениях, а в последнее время и учение о радикале оказались перенесенными в общую теорию алгебраических систем. Это позволяет говорить об этой теории как о новом отделе алгебры.

Рассматривая характер алгебраической науки в целом, часто подчеркивают как отличительную ее особенность отсутствие или подчиненность понятия непрерывности, признавая тем самым алгебру наукой по преимуществу о дискретном. Такой взгляд, несомненно, отражает одну из важных объективных особенностей алгебры. В реальном мире прерывное и непрерывное находится в диалектическом единстве. Но чтобы познать действительность, иногда необходимо ее рассечь на части и изучать эти части порознь. Поэтому одностороннее внимание алгебры к дискретным соотношениям нельзя рассматривать как ее недостаток.

На примере теорий групп видно, что отдельные алгебраические дисциплины дают не только средства для технических вычислений, но и язык для выражения глубоких законов природы. Однако, помимо непосредственного практического значения ряда разделов алгебры для физики, химии, кристаллографии и других наук, в самой математике алгебра занимает одно из важных мест. По словам замечательного советского алгебраиста Н. Г. Чеботарева, алгебра была колыбелью многих новых идей и понятий, возникших в математике, и в значительной степени оплодотворяла развитие таких разделов математики, которые служат уже непосредственной базой физических и технических наук.

ЛИТЕРАТУРА[1]

Александров П. С. Введение в теорию групп. Учпедгиз, 1952.
Ван дер Варден. Современная алгебра, ч. I и II. Гостехиздат, 1947.
Джекобсон Н. Теория колец. ИЛ, 1947.
Курош А. Г. Теория групп. Гостехиздат, 1953.
Понтрягин Л. С. Непрерывные группы. Гостехиздат, 1954.

Краткие исторические сведения и обсуждение основных методологических вопросов содержатся в БСЭ, в статьях: «Алгебра», «Групп теория», «Математика» и др. Обзор достижений советских математиков в области алгебры до 1947 г. помещен в сборнике «Математика в СССР за тридцать лет», Гостехиздат, 1948.

Литература по федоровским группам указана на стр. 275.

[1] Общие курсы алгебры были указаны в литературе к главам IV и XVI.

ИМЕННОЙ УКАЗАТЕЛЬ

СОДЕРЖАНИЕ ПЕРВОГО И ВТОРОГО ТОМОВ

ТОМ I

ТОМ II

ОГЛАВЛЕНИЕ

Математика, ее содержание, методы и значение
Том III

*

Утверждено к печати Математическим институтом им. В. А. Стеклова Академии наук СССР

*

Редактор издательства *А. З. Рывкин*. Технический редактор *Е. В. Зеленкова*

РИСО АН СССР № 28,13В. Сдано в набор 20/VIII 1956 г. Подписано к печати 4/XII 1956 г. Формат 70×108 1/16. Печ. л. 21. Усл. л. 28,77. Уч.-изд. л. 23. Тираж 7000 экз. Т-11726. Изд. № 1592. Тип. зак. № 812

Цена 17 руб. 60 коп.

*

Издательство Академии наук СССР. Москва Б-64. Подсосенский пер., 21

1-я типография Издательства АН СССР. Ленинград, В. О., 9 линия, 12.