MATHEMATICS AND LOGIC

Retrospect and Prospects

MARK KAC AND STANISLAW M. ULAM

FREDERICK A. PRAEGER, PUBLISHERS
New York • Washington • London

1968

«СОВРЕМЕННАЯ МАТЕМАТИКА»

Популярная серия

М. КАЦ, С. УЛАМ

Математика и логика

РЕТРОСПЕКТИВА И ПЕРСПЕКТИВЫ

Перевод с английского
Н. И. Плужниковой

Под редакцией
И. М. Яглома

ИЗДАТЕЛЬСТВО «МИР»

Москва 1971

УДК 517.1

Книга видных американских ученых Марка Каца и Станислава Улама (оба автора хорошо известны советскому читателю по переводу ряда других их книг и статей) была подготовлена для выпускаемой издательством Британской энциклопедии серии обзоров, посвященных состоянию и ближайшим перспективам развития различных наук. Рассчитанная на широкий круг читателей, книга ставит своей целью освещение современного состояния математики и ее специфических черт. Особое место уделяется взаимодействию и взаимозависимости математики и других наук, обогащающих, по мнению авторов, как чистую математику, так и все использующие математические методы направления научной мысли, а также обсуждению возможного будущего математики.

Интересная по содержанию и блестящая по форме книга М. Каца и С. Улама бесспорно привлечет внимание читателей самых разных кругов.

Редакция литературы по математическим наукам

Инд. 2-2-3
34-71

ОТ ИЗДАТЕЛЬСТВА

Роль математики в научно-техническом прогрессе весьма велика, а в последние десятилетия она возросла особенно. Математика, которая традиционно обслуживала механику, астрономию и некоторые разделы физики, сейчас активно вторгается в технику, экономику, науку об управлении, начинает завоевывать плацдармы в медицине, биологии и других областях естественных и общественных наук. Возрастает и интерес к математике среди неспециалистов, особенно молодежи, желание разобраться в особенностях математических методов, понять, в чем сила и привлекательность этой науки.

Книга написана по заказу Британской энциклопедии и призвана дать представление о математике, доступное достаточно широким читательским кругам. Ее авторы — американские ученые, выходцы из Польши. Оба они известны не только научными результатами, но и своей деятельностью по популяризации науки. Марк Кац в течение ряда лет руководил Исследовательской группой по школьной математике США, Станиславу Уламу принадлежит собрание задач и проблем из разных разделов математики. Советским читателям М. Кац и С. Улам известны по переводу ряда их книг и статей.

Большую часть книги занимает первая глава, которая называется «Примеры». В ней разбирается много конкретных задач из разных областей математики, иллюстрирующих богатство и своеобразие ее идей. Задачи эти составляют, так сказать, экспериментальный фактический материал, без которого понять, что представляет собой математика, невозможно. Они подобраны с большим вкусом, изложены живо, интересно и по возможности доступно. Авторы начинают с широко известных идей расширения понятия числа и осуществимости геометрических построений, затем переходят к основам теории вероятностей и, наконец, к теории групп и линейной алгебре. По мере накопления фактов обсуждаются их связи друг с другом и с физикой — постепенно из отдельных деталей начинает вырисовываться величественная картина математики.

Следующая глава «Темы, тенденции и синтез» призвана завершить создание этой картины. Здесь оттеняются наиболее важные идеи, поданные в их развитии. Особое место уделяется логике и основаниям математики, а также изменениям, которые вызвало в математике появление электронных вычислительных машин.

Третья глава посвящена связям математики с другими науками. Здесь, в частности, коротко излагаются задачи массового обслуживания (теории очередей), теории игр и теории информации. В небольшой последней главе «Итоги и перспективы» уже совсем бегло описываются некоторые самые свежие исследования и делаются предположения о дальнейшем их развитии.

Не со всеми утверждениями авторов можно согласиться. В частности, вызывает возражения неоднократно высказываемая ими мысль о том, что внешний мир является лишь источником математических понятий и теорий, а дальше математика развивается независимо по своим внутренним законам. Конечно же, связи математики с внешним миром неизмеримо глубже и богаче. Чтобы составить правильное представление об этом и других методологических вопросах, читатель должен обратиться, например, к статье А. Н. Колмогорова «Математика» в 26-м томе Большой советской энциклопедии и к книге «Математика, ее содержание, методы и значение» (Изд-во АН СССР, М., 1956).

Спорным является и то, что авторы выделяют из математики и относят к другим наукам такие важные разделы, как теория игр и теория информации. Пожалуй, слишком критически они относятся и к возможностям применения математических методов в экономике и совсем ничего не говорят о применениях в психологии, педагогике и других науках, изучающих интеллектуальную деятельность.

С сожалением приходится отметить, что вообще новые разделы математики, особенно прикладные, не нашли в книге столь полного и яркого отражения, как классические. По-видимому, эти разделы заслуживают отдельной книги того же объема и столь же интересной. Но такая книга еще не написана...

Эта же книга, несмотря на отмеченные недостатки, доставит истинное удовольствие всем, кто любит математику, независимо от того, знает он ее, или только начинает с ней знакомиться.

ВВЕДЕНИЕ

Что такое математика? Как она возникла, кто ее создал и кто делает ее сейчас? Можно ли обрисовать путь ее развития и ее место в истории научной мысли? Можно ли предсказать ее будущее? Эта книга — попытка разобраться в подобных вопросах и дать читателю представлеиие о необъятиости и глубине предмета.

Математика — это замкнутый в себе микрокосм, обладающий, однако, мощной способностью отражать и моделировать любые процессы мышления и, вероятно, всю науку вообще. Она всегда приносила большую пользу и еще в большей мере продолжает приносить ее сейчас. Можно даже пойти дальше и сказать, что математика необходима для покорения природы человеком и вообще для развития человека как биологического вида, ибо она формирует его мышление.

В самом деле, иасколько можно проследить по летописям человеческой любознательности и стремления к знанию, математику всегда бережно лелеяли, заботясь о передаче накопленных истин новым поколениям. Ее рассматривали как окончательное выражение рациональных размышлений о внешнем мире и как памятник извечному желанию человека испытать работу своего ума. Мы не будем пытаться дать *определение* математики, так как для этого пришлось бы втиснуть ее в какие-то границы. Как увидит читатель, математика способна обобщить любую схему, изменить и обогатить ее. И все же каждый раз, когда это происходит, полученный результат по-прежнему составляет лишь часть математики. Вероятно, самым характерным для этой науки является то, что она развивается путем постоянной самопроверки и все более глубокого осознания собственной структуры. Сама же эта структура непрерывно изменяется, причем иногда весьма радикально. Поэтому любая попытка дать сколько-нибудь полное и исчерпывающее определение математики обречена, по нашему мнению, на неудачу.

Мы попробуем описать в историческом аспекте ряд наиболее важных моментов развития математической мысли. При этом мы нередко будем задерживаться на вопросе о том, в какой мере прогресс в математике связан с «придумыванием» и до какой степени он носит характер «открытия». Иначе говоря, вопрос состоит в том, диктуется ли выбор аксиом, определений и проблем окружающим нас внешним миром, который мы воспринимаем своими органами чувств, наблюдаем и измеряем при помощи разных приборов и инструментов, или же эти аксиомы, определения и проблемы являются свободными творениями человеческого разума и определяются физиологической структурой мозга?

За последние пятьдесят лет математика подобно другим наукам претерпела большие изменения. И дело не только в том, что значительно вырос объем математических знаний и изменились представления о важности тех или иных проблем. Другими в какой-то мере стали тон и цели математики. Многие великие успехи физики, астрономии и других «точных» наук были достигнуты главным образом благодаря математике. Свободно заимствуя средства, предоставляемые математикой, эти родственные науки в свою очередь снабжают ее новыми проблемами и открывают новые источники вдохновения.

Глубокое влияние на математику оказывает и развитие техники: так, создание быстродействующих вычислительных машин неизмеримо расширило возможности экспериментирования в самой математике.

Коренные изменения коснулись и оснований математики и математической логики. В главе 2 мы постараемся объяснить, в чем *сущность* этих изменений.

В процессе своего развития математика постоянно возвращается к некоторым специфическим понятиям и темам; мы приведем много примеров, иллюстрирующих их вариации и взаимодействие.

Одним из таких понятий, наиболее характерных для математики, является *бесконечность*. Мы уделили этой теме довольно много места, пытаясь показать, как возникло представление о бесконечности и как оно затем развивалось и уточнялось.

Вопреки мнению, широко распространенному среди далеких от науки людей, математика не являет собой законченное и совершенное здание. Это наука и, кроме того, искусство; свойственные математике критерии всегда в той или иной степени носят эстетический характер. Одной только истинности теоремы еще недостаточно для того, чтобы считать ее достойной занять место в математике. Она должна быть «полезной», «интересной» и «красивой». А поскольку ощущение красоты субъективно, остается только удивляться, какое единодушие проявляют обычно математики в своих эстетических оценках.

В одном отношении математика стоит особняком среди других наук: никакой ее результат не может быть зачеркнут дальнейшим развитием науки. Однажды доказанная теорема уже никогда не станет неверной, хотя впоследствии может выясниться, что она является лишь тривиальным частным случаем какой-то более общей истины. Математические знания не подлежат пересмотру, и общий их запас может лишь возрастать.

Можно ли при таком невероятном разнообразии проблем и способов применения усмотреть в математике какой-то внутренний порядок? Что обеспечивает ей не вызывающее сомнений единство и что делает ее самостоятельной наукой?

Начнем с того, что следует различать *объекты* математики и ее *метод*.

Самыми первыми математическими объектами являются натуральные числа 1, 2, 3, ..., а также точки и простые геометрические фигуры (прямые линии, треугольники и т. п.). Они настолько привычны и знакомы нам с детства, что долгое время считались не требующими никаких пояснений и уточнений. Только в конце прошлого века был впервые предпринят серьезный логический анализ арифметики (Пеано, Фреге, Рассел) и геометрии (Гильберт). Однако столь характерный для математики процесс образования новых объектов и выделения новых структур происходил и до того, как были уточнены понятия натурального числа и точки.

От объектов можно перейти к *множествам* этих объектов, к *функциям* и *соответствиям*. (Идея соответствия, или преобразования, родилась из естественного стремления человека распознать родственные друг другу расположения и выделить общую *закономерность*, объединяющую различные на вид ситуации.) Продолжая возникающий здесь процесс *итерации*, переходят к классам функций, к соответствиям между функциями (операторам), затем к классам операторов и т. д.; этот процесс, шаги которого становятся все крупнее, не имеет конца. Таким способом из простых объектов получают все более и более сложные.

Главной составной частью математического метода является *доказательство*, структура которого едва ли сильно изменилась со времен греков. По-прежнему сначала постулируется небольшое число *аксиом* (предложений, которые принимаются на веру), а затем по строгим логическим правилам выводятся новые предложения. Сам этот процесс, его сильные и слабые стороны — все это лишь в сравнительно недавние годы стало предметом критического изучения. Метаматематика, которая этим занимается, сама является частью математики. *Объектом* метаматематики служит довольно ограниченный, казалось бы, набор правил, относящихся к математической логике. Однако насколько всеобъемлющи и всемогущи эти правила! В какой-то мере математика питает сама себя, но здесь нет порочного круга: как показывают блестящие успехи математического метода в физике, астрономии и других естественных науках, — это отнюдь не бесплодная игра. Вероятно, так происходит потому, что многие объекты математического исследования навеяны внешним миром, а обобщение и выбор новых структур тоже не совсем произвольны. «Непостижимая эффективность математики», быть может, и остается чудом для философии, однако это никак не отражается на ее очевидных и несомненных успехах.

Математику определяли как науку вывода необходимых следствий. Но каких именно следствий? Простая цепь силлогизмов — это еще не математика. Каким-то образом мы выбираем те формулировки, которые в компактном виде охватывают широкий класс частных случаев, а определенные доказательства считаем элегантными или красивыми. Значит, метод включает в себя нечто большее, чем простая логика, заключенная в дедукции. Объекты же содержат в себе меньше, чем их интуитивные или инстинктивные источники.

Отличительная черта математики — возможность оперировать объектами, не определяя их.

Точки, прямые, плоскости *не определяются.* В наши дни математик отвергает попытки своих предшественников определить точку как нечто, не имеющее «ни длины, ни ширины», или дать столь же бессмысленные псевдоопределения прямой и плоскости.

В течение многих веков выработалась такая точка зрения: неважно, что именно *представляют собой* рассматриваемые объекты, если известно, какие *утверждения* о них *допустимы.* Знаменитая работа Гильберта «Основания геометрии» начинается словами: «Пусть имеется три типа объектов; объекты первого типа будем называть «точками», второго типа — «прямыми», третьего — «плоскостями»». Вот и все, — а за этим следует перечень исходных утверждений (аксиом), включающих слова «точка», «прямая» и «плоскость»; из этих аксиом можно выводить, пользуясь уже только правилами логики, дальнейшие утверждения, содержащие эти не определенные нами слова. Такой подход позволяет научить геометрии слепого и даже вычислительную машину!

Этот характерный тип абстракции, ведущий к полному игнорироваиию физической природы геометрических объектов, встречается не только в традиционных границах математики. Примером служит предложенная Эрнстом Махом (на основе работ Джеймса Максвелла) трактовка понятия температуры. Для определения температуры необходимо ввести понятия *теплового равновесия* и *теплового контакта*; последние же крайне трудно, если вообще возможно, определить в логически приемлемых терминах. Однако, как показывает анализ, все, что на самом деле нужно, — это свойство *транзитивности* теплового равновесия, т. е. постулат (называемый иногда *нулевым началом* термодинамики), утверждающий, что если (A и B) и (A и C) находятся в тепловом равновесии, то и (B и C) находятся в тепловом равновесии. [Для полноты нужно еще добавить в некотором смысле обращение этого нулевого начала: если A, B и C находятся в тепловом равновесии, то этим свойством обладают (A и B) и (A и C).] Так же, как в геометрии, здесь необязательно знать логически точный смысл терминов; достаточно уметь объединять их в осмысленные (т. е. допустимые) предложения.

Хотя мы можем успешно оперировать с неопределенными (и, вероятно, даже неопределимыми) объектами и понятиями, сами эти объекты и понятия уходят своими корнями в видимый физический (или по крайней мере чувственный) мир. Физические явления подсказывают и даже диктуют нам исходные аксиомы, и под влиянием той же видимой физической реальности мы формулируем вопросы и проблемы.

Анри Пуанкаре сказал, что существовать в математике — значит быть свободным от противоречий. Но одно только существование еще не гарантирует возможности выжить. Чтобы выжить в математике, нужна такая разновидность жизнеспособности, которую не опишешь в чисто логических терминах.

В следующих главах мы обсудим ряд проблем, которые не только проявили завидную жизнестойкость, но и положили начало развитию наиболее плодотворных математических теорий. Они простираются от конкретного к абстрактному и от самого простого к сравнительно сложному. Мы выбрали их, чтобы проиллюстрировать объекты математики и математический метод и убедить читателя, что чистая математика все-таки не исчерпывается определеиием Рассела: «Чистая математика есть класс всех предложений вида «из p следует q», где p и q — предложения, содержащие одиу или более перемеиных, одни и те же в обоих предложениях, и ни p, ни q не содержат никаких постояиных, кроме логических постояииых».

ГЛАВА 1

ПРИМЕРЫ

§ 1. Бесконечность множества простых чисел

Среди так называемых *натуральных чисел* (1, 2, 3 и т. д.) имеются такие, которые делятся только на 1 и на себя; они называются *простыми числами*. Простые числа — это «кирпичи», из которых строятся все остальные натуральные числа: каждое натуральное число есть произведение степеней своих простых делителей; например, $60=2^2\cdot3\cdot5$. Вот несколько первых простых чисел: 1, 2, 3, 5, 7, 11, 13, 17. Возникает вопрос: можно ли продолжать эту последовательность бесконечно? Иначе говоря: существует ли самое большое простое число? Ответ известен еще со времен античной Греции: наибольшего простого числа не существует. Это доказал Евклид в третьем веке до нашей эры, и сегодня его рассуждение так же свежо и прозрачно, как и в древности. Когда было установлено, что простых чисел бесконечно много, сразу возникло множество других вопросов об этих числах; некоторые из них — это известные всем математикам проблемы, до сих пор остающиеся нерешенными. В этом параграфе мы обсудим несколько таких вопросов и приведем принадлежащее Евклиду доказательство неограниченности последовательности простых чисел.

Мы не знаем, когда впервые появилось понятие простого числа и сколько времени протекло от установления простейших свойств таких чисел до открытия, что их бесконечно много. Вопрос об их количестве возник, вероятно, вскоре после первых попыток «экспериментального» и «прагматического» изучения таких чисел, как 2, 3, 11, 17. Идея бесконечности, которой, по-видимому, предшествовало понятие «произвольно большого», должна была формироваться значительно дольше; возможно, она возникла в результате наблюдения окружающего физического мира.

Следующее доказательство — вероятно, до сих пор самое простое — устанавливает только существование произвольно больших простых чисел. Предположим, что множество простых чисел конечно; тогда существует самое большое простое число, скажем p. Рассмотрим число $n = p! + 1$ ($p!$ читается «пе факториал» и равно произведению $1 \cdot 2 \cdot 3 \ldots p$). Оно не делится ни на какое простое число вплоть до p. Поэтому либо между p и n должно быть какое-нибудь простое число, либо простым является само n. И то, и другое противоречит нашему предположению, что p — наибольшее простое число.

Этот удивительно простой и изящный результат Евклида — один из первых известных нам доказательств от противного. Как свойственно всей «хорошей» математике, он разрешает некоторый вопрос, подсказывает новые вопросы и приводит к новым наблюдениям. Например, снова используя факториалы, мы можем немедленно установить, что в последовательности натуральных чисел имеются сколь угодно длинные «куски», не содержащие простых чисел. В самом деле, для любого заданного n можно выписать $n - 1$ последовательных чисел: $n! + 2$, $n! + 3$, ..., $n! + n$, первое из которых делится на 2, второе — на 3, и т. д. (последнее делится на n).

Остановимся еще немного на этом примере, так как он хорошо иллюстрирует характерную особенность математического мышления: постановку все новых и новых проблем, неизбежно приводящих к новым, более трудным, а иногда и неразрешимым.

Уверенные теперь в том, что последовательность простых чисел бесконечна, мы хотим знать больше. Можно ли найти частоту простых чисел или оценить количество $\pi(n)$ простых чисел между 1 и некоторым (большим) целым n? Можно доказать, что $\pi(n)$ асимптотически равно $n/\log n$ [1]), т. е. что отношение

[1]) Здесь и везде далее речь идет о натуральных логарифмах, т. е. о логарифмах при основании e. — *Прим. ред.*

$\pi(n)$ к $n/\log n$ приближается к 1 при неограниченном росте n. Эта замечательная теорема о простых числах была впервые доказана в 1896 г. Ж. Адамаром и Ш.-Ж. Валле Пуссеном. Первые доказательства включали довольно сложные понятия математического анализа: они опирались на результаты теории аналитических функций. Только значительно позднее П. Эрдёш и А. Сельберг нашли более элементарное (хотя длинное и сложное) доказательство, использующее лишь комбинаторные и арифметические понятия и не требующее никаких сведений об аналитических функциях.

Все предыдущее относилось к *количеству* простых чисел по сравнению со всеми целыми числами. Другие наблюдения над простыми числами немедленно возбуждают самое элементарное любопытство. Так, например, какое бы мы ни взяли конкретное четное число, его можно представить в виде суммы двух простых чисел. Математик X. Гольдбах (1690—1764) выдвинул гипотезу, что это верно для всех четных чисел. Его гипотеза до сих пор остается недоказанной; проверено лишь, что она выполняется для четных чисел вплоть до 100 000 000. Используя электронные вычислительные машины, можно даже собрать статистические данные, показывающие, сколькими различными способами то или иное четное число $2n$ разбивается в сумму двух простых; оказывается, число способов довольно быстро растет с ростом n. Уже в наши дни советский математик И. М. Виноградов доказал, что любое достаточно большое нечетное число можно представить в виде суммы трех простых чисел!

Не существует никакой формулы или выражения, позволяющего *записать* сколь угодно большое простое число. Известны некоторые арифметические выражения, дающие много простых чисел; например, формула Эйлера $N = x^2 + x + 41$ приводит к простым числам при $x = 0, 1, 2, \ldots, 39$, однако мы не знаем, имеется ли бесконечно много целых x, таких, что $x^2 + x + 41$ — простое число. Существуют ли вообще многочлены от x, принимающие при подстановке в них целых x бесконечно много простых значений?

Известно только, что такие многочлены первой степени существуют (таков, например, многочлен $2x+1$), однако до сих пор неизвестно, обладает ли этим свойством какой-нибудь многочлен степени больше 1. Неизвестно также, существует ли бесконечно много «близнецов», т. е. простых чисел, различающихся на 2 (например, 11 и 13, 29 и 31 и т. д.).

Примеры такого рода типичны для любой ветви математики: вопросы, возникающие почти автоматически, часто оказываются чрезвычайно трудными, хотя их формулировка не дает оснований предполагать, что они выходят далеко за пределы уже установленных фактов.

Тем не менее о последовательности простых чисел известно очень многое сверх того, о чем сказано выше. Например, доказано, что существует бесконечно много простых чисел вида $4k+1$ и бесконечно много простых чисел вида $4k+3$. Более того, известно, что любая арифметическая прогрессия $a \cdot k + b$ (где a и b — взаимно простые целые числа, а $k = 1, 2, 3, \ldots$) содержит бесконечно много простых чисел и что эти простые числа асимптотически имеют «правильную» частоту (т. е. частоту $1/\varphi(a)$[1]).

Приведенные примеры и комментарии к ним должны проиллюстрировать одну нить математического мышления, которая, разветвляясь, тянется через всю его историю: математики, обнаружив какие-нибудь свойства конечной совокупности чисел (первых и самых важных математических объектов), стремятся доказать эти свойства для бесконечного множества чисел или всей их совокупности. Эта тяга математиков к обобщениям всегда была самой характерной чертой математики. Слова Л. Кронекера: «целые числа созданы богом, а все остальное в математике — дело рук человеческих» являются крайним выражением этой точки зрения. Впрочем, с ними можно и не

[1]) Это означает, что если $\pi_a(n)$ — количество простых чисел в прогрессии $a \cdot k + b$, не превосходящих n, то $\lim\limits_{n\to\infty} [\pi_a(n)/\pi(n)] = 1/\varphi(a)$, где $\varphi(a)$ — количество целых чисел, меньших a и взаимно простых с a.

соглашаться, поскольку простые геометрические объекты, несомненно, имеют не меньшее право претендовать на первозданность.

§ 2. Иррациональность числа $\sqrt{2}$

В системе так называемых *натуральных чисел* 1, 2, 3, ... не всегда возможно вычитание: $3-5=-2$ — не натуральное число, поскольку оно отрицательно. В расширенной системе *целых чисел* $0, \pm 1, \pm 2, \ldots$ не всегда возможно деление: $2:3=2/3$ — не целое число. В еще более обширной системе всех дробей (*рациональных чисел*) деление (за исключением деления на 0) всегда выполнимо. В этом параграфе будет показано, что множество рациональных чисел еще недостаточно «богато» для всех целей, которые могут возникнуть в арифметике: в системе рациональных чисел не всегда возможно извлечение корня. Например, квадратные корни из некоторых рациональных чисел не являются рациональными. Мы покажем, что $\sqrt{2}$ — не рациональное число. Изучение таких «иррациональностей», начатое еще во времена древней Греции, имело большое значение для дальнейшего развития математики. Оно привело к новым важным понятиям и результатам; некоторые из них мы обсудим в этом параграфе.

Иррациональность можно описать при помощи представления числа в виде десятичной дроби. Если дробь конечная (как $1/4=0{,}25$) или периодическая (как $1/3=0{,}333\ldots$), рассматриваемое число есть отношение целых чисел и, следовательно, рационально; в противном случае оно называется *иррациональным*. Разумеется, для того чтобы сказать, является ли соответствующая числу десятичная дробь конечной или периодической, нужно знать полностью эту бесконечную десятичную дробь; ясно, что это невозможно. Поэтому приходится применять другие методы, подобные тому, который мы используем ниже.

Иррациональные числа бывают двух типов. Одни являются корнями алгебраических уравнений (например, $\sqrt{2}$ есть корень уравнения $x^2-2=0$) и поэтому называются *алгебраическими* числами. Число, не являющееся корнем никакого алгебраического уравнения с рациональными коэффициентами, называется *трансцендентным*, поскольку оно «переступает пределы» операций обычиой арифметики. Примерами трансцендентных чисел служат число π и основание системы натуральных логарифмов e. Вопрос о том, как для некоторого числа определить, является ли оно рациональным, алгебраическим или трансцендентным, в общем виде еще не решен.

Все рациональные и иррациональные числа вместе образуют систему *действительных чисел*. Для того чтобы были разрешимы все алгебраические уравнения (в частности, уравнение $x^2+1=0$), необходимо расширить снстему действительных чисел присоединением так называемой «мнимой единицы» $i=\sqrt{-1}$. Множество всех чисел вида $a+ib$ с действительными a и b называется системой *комплексных чисел*. Подробнее вопрос о числовых системах мы рассмотрим в § 8.

Большую роль в развитии математики всегда играли доказательства *невозможности* определенных построений, ограничивающие поле действия определенных теорий и методов. Это многократно приводило к обогащению существующих математических понятий, к расширению систем аксиом и к введению новых объектов. Мы попытаемся проиллюстрировать это на, видимо, самом первом примере такого рода: доказательстве иррациональиости числа $\sqrt{2}$.

Можно ли выразить длину диагонали квадрата со стороной 1 в виде отношения двух целых чисел? Иначе говоря, существуют ли взаимно простые целые числа a и b, такие, что $(a/b)^2=2$? Вот данное греками замечательно простое доказательство того, что это не так. Если $a^2=2b^2$, то a четно, поскольку $2b^2$ четно, а квадрат нечетного числа не может быть четным. Тогда $a=2a_1$, где a_1 — снова целое. Следовательно, $a^2=4a_1{}^2=2b^2$; разделив последнее равенство на 2, мы заключим, что b тоже четно: $b=2b_1$. Но мы предположили, что дробь a/b несократима, т. е. a и b не имеют общих множителей. Это противоречие доказывает *невозможность* представления $\sqrt{2}$ в виде дроби; таким образом, $\sqrt{2}$ — это не дробь, а число какого-то другого вида, которому мы еще должны дать *определение*.

Хотя иррациональность $\sqrt{2}$ давно известна, это утверждение можно перефразировать столь удивительным образом, что оно кажется почти парадоксальным.

Рассмотрим все рациональные числа в интервале от 0 до 1, за исключением 0. Каждое из них можно единственным образом представить в виде дроби a/b, где a и b не имеют общих множителей. Допустим, что a/b — середина интервала длины $1/(2b^2)$; иначе говоря, *покроем* a/b *интервалом с концами* $a/b - 1/(b^2)$ *и* $a/b + 1/(b^2)$. Рациональные числа образуют всюду плотное множество (т. е. в любом сколь угодно малом интервале всегда имеются рациональные числа), а сумма длин покрывающих интервалов бесконечна; поэтому при таком покрытии рациональных чисел мы должны, казалось бы, автоматически покрыть и все остальные числа. Однако число $\sqrt{2}/2$ осталось непокрытым! В самом деле, число[1] $|b^2-2a^2|$, будучи целым, должно быть не меньше 1; оно не может равняться нулю ввиду иррациональности $\sqrt{2}$. Таким образом,

$$\frac{|b^2-2a^2|}{2b^2} \geqslant \frac{1}{2b^2},$$

$$\left|\frac{\sqrt{2}}{2}-\frac{a}{b}\right|\left(\frac{\sqrt{2}}{2}+\frac{a}{b}\right) \geqslant \frac{1}{2b^2},$$

$$\left|\frac{\sqrt{2}}{2}-\frac{a}{b}\right| \geqslant \frac{1}{2b^2}\frac{1}{\frac{\sqrt{2}}{2}+\frac{a}{b}} > \frac{1}{2b^2}\frac{1}{2} = \frac{1}{4b^2},$$

откуда следует, что число $\sqrt{2}/2$ не покрыто.

---

[1]) Как обычно, $|x|$ означает абсолютную величину x; так, $|-5| = 5$, $|2| = 2$ и т. д.

Доказательство иррациональности $\sqrt{2}$ (с подходящими изменениями) можно использовать для того, чтобы убедиться в иррациональности $\sqrt{5}$, и т. д. Квадратичные иррациональности можно получать при помощи геометрических построений, однако уже греки не ограничились в своих размышлениях лишь иррациональностями этого типа: одной из их самых старых задач была задача построения отрезка длины $\sqrt[3]{2}$ (так называемая делосская задача об удвоении куба). Прошли века, прежде чем была доказана невозможность построения $\sqrt[3]{2}$ циркулем и линейкой.

И опять, как и в нашем первом примере, сразу же возникают новые вопросы. Существуют ли действительные числа, которые нельзя получить как корни алгебраического уравнения с целыми коэффициентами? Лишь в 19-м веке впервые была доказана трансцендентность чисел e и π.

Один способ построения трансцендентных чисел был указан Ж. Лиувиллем. И только после того, как Кантор разработал теорию множеств, удалось установить, что «большинство» действительных чисел не являются алгебраическими. Оказалось, что алгебраические числа образуют лишь счетное множество, т. е. их можно выписать в одну последовательность a_1, a_2, a_3, ..., в то время как совокупность всех действительных чисел нельзя расположить таким способом. Доказательство этого утверждения читатель найдет в § 4.

Существуют числа, для которых до сих пор не установлено, рациональны они или нет, несмотря на то, что легко описать, как они получены. Одним из таких чисел является постоянная Эйлера, определяемая следующим образом. Рассмотрим ряд $\frac{1}{2}+\frac{1}{3}+\frac{1}{4}+\ldots+\frac{1}{n}+\ldots$; n-я частная сумма (т. е. сумма первых n членов) этого ряда близка к $\log n$. Разность между этой суммой и $\log n$ при возрастании n стремится к некоторому пределу; его называют постоянной Эйлера и обозначают буквой C. Известно, что $C \approx 0{,}6$. Несмотря на старания многих математиков, вопрос о том, рационально или нет число C, остается открытым; вполне возможно, что оно даже не алгебраическое!

Квадратичные иррациональности разлагаются в непрерывные дроби весьма специального вида. Всякое действительное число можно записать в виде

$$a_0+\cfrac{1}{a_1+\cfrac{1}{a_2+\cfrac{1}{a_3+\ldots}}},$$

где все a_i — целые. Для квадратичных иррациональностей (например, для числа $\sqrt{2}$ и вообще для всех чисел вида $p+q\sqrt{r}$, где p, q, r рациональны) эти a_i образуют периодическую последовательность; обратно, если последовательность a_0, a_1, a_2, ... — периодическая, то выражаемое соответствующей непрерывной дробью число x представляет собой квадратичную иррациональность. Поэтому в разложении заданного числа x, являющегося квадратичной иррациональностью, все значения a_i ограничены, т. е. не превосходят некоторого фиксированного числа (зависящего от x). Неизвестно, существует ли хотя бы одно иррациональное алгебраическое число порядка выше 2 (например, кубическая иррациональность или иррациональность еще более высокого порядка), для которого все числа a_i были бы ограничены. Недавно была установлена трансцендентность некоторых чисел (например, $\sqrt{2}^{\sqrt{2}}$).

§ 3. Приближения рациональными числами

Иррациональные числа обычно приближают рациональными. Например, $\sqrt{2}$ приближенно равно $1{,}4 = 140/100$, а π приближенно равно $3{,}14 = 314/100$. Каждое действительное число можно сколь угодно точно приблизить рациональными числами, приписывая, например, все новые и новые цифры в конце соответствующей бесконечной десятичной дроби. Однако уже самые элементарные приемы позволяют получить и значительно более точную и полную информацию о методах приближения; этим мы и займемся в настоящем параграфе. Используя полученную информацию, мы построим затем одно трансцендентное число.

Многие глубокие и элегантные математические исследования посвящены вопросам *приближения* алгебраических и других чисел рациональными дробями. Насколько хорошо можно приблизить, скажем, $\sqrt{2}$ дробью a/b? Это можно сделать с любой точностью — ведь всякое действительное число является пределом некоторой последовательности рациональных чисел. Однако хотелось бы сделать это достаточно «экономно», т. е. добиться того, чтобы величина $\varepsilon = |\sqrt{2} - a/b|$ была как можно меньше при возможно меньшем b. Оказывается, всегда можно сделать ε меньшим, чем c/b^2, где c — некоторая постоянная; при этом было выяснено, что из всех чисел труднее всего приблизить дробями с точностью c/b^2 квадратичные иррациональности. Среди таких наиболее трудных для приближения чисел можно указать $\frac{\sqrt{5}-1}{2}$ — число, которое греки называли «золотым», поскольку оно получается при делении отрезка в крайнем и среднем отношении (так называемое «золотое сечение» отрезка); для этого числа значение c является наибольшим.

В то же время некоторые трансцендентные числа допускают очень хорошее приближение в указанном смысле; этот факт был использован Лиувиллем при построении трансцендентного числа, носящего его имя. Лиувилль первым доказал следующую теорему:

Если α — корень неприводимого алгебраического уравнения с целыми коэффициентами степени $n \geqslant 2$, то существует положительная постоянная γ, зависящая *только* от α и такая, что для *всех* целых p, q

$$\left|\alpha - \frac{p}{q}\right| > \frac{\gamma}{q^n} \quad (\gamma > 0).$$

Доказательство теоремы Лиувилля достаточно простое, но в нем используются элементы дифференциального исчисления. Пусть

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

— неприводимый многочлен с целыми коэффициентами ($a_0 \neq 0$, так как степень уравнения равна n), одним из корней которого является α. Производная $f'(x)$ на отрезке $[\alpha - 1, \alpha + 1]$ ограничена, т. е. существует такое число M, что

$$|f'(x)| \leqslant M \quad \text{при} \quad \alpha - 1 \leqslant x \leqslant \alpha + 1.$$

Достаточно рассмотреть лишь те рациональные числа p/q, которые лежат в интервале между $\alpha - 1$ и $\alpha + 1$. Далее,

$$\left|f\left(\frac{p}{q}\right)\right| = \frac{|a_0p^n + a_1p^{n-1}q + \ldots|}{q^n} \geqslant \frac{1}{q^n},$$

поскольку $f\left(\frac{p}{q}\right) \neq 0$ (многочлен неприводим) и $|a_0p^n + a_1p^{n-1}q + \ldots|$ — целое число.

Используя теорему о среднем из дифференциального исчисления, мы заключаем, что между α и p/q (и, следовательно, в интервале от $\alpha - 1$ до $\alpha + 1$) найдется такое число x, что

$$f(\alpha) - f(p/q) = (\alpha - p/q)\, f'(x),$$

откуда, поскольку $f(\alpha) = 0$,

$$\frac{1}{q^n} \leqslant \left|f\left(\frac{p}{q}\right)\right| = \left|f(\alpha) - f\left(\frac{p}{q}\right)\right| = \left|\alpha - \frac{p}{q}\right| |f'(x)| \leqslant M\left|\alpha - \frac{p}{q}\right|,$$

или

$$\left|\alpha - \frac{p}{q}\right| \geqslant \frac{1}{M}\frac{1}{q^n},$$

что и требовалось доказать.

Рассмотрим теперь число

$$\alpha = \frac{1}{10} + \frac{1}{10^{2!}} + \frac{1}{10^{3!}} + \frac{1}{10^{4!}} + \ldots =$$
$$= \frac{1}{10} + \frac{1}{10^2} + \frac{1}{10^6} + \frac{1}{10^{24}} + \ldots =$$
$$= 0{,}110\,001\,000\,000\,000\,000\,000\,001\,0\ldots$$

Можно проверить, что

$$0 < \alpha - \left(\frac{1}{10} + \frac{1}{10^{2!}} + \ldots + \frac{1}{10^{m!}}\right) < \frac{2}{10^{(m+1)!}},$$

т. е. что существует последовательность целых чисел p_m, для которых

$$0 < \alpha - \frac{p_m}{10^{m!}} \leqslant 2\left(\frac{1}{10^{m!}}\right)^{m+1}.$$

Иными словами, существует последовательность рациональных чисел p_m/q_m, где $q_m = 10^{m!}$, таких, что

$$0 < \alpha - \frac{p_m}{q_m} \leqslant \frac{2}{q_m^{m+1}}.$$

Если бы α было алгебраическим, то при некотором *фиксированном* n для *всех* m выполнялось бы неравенство

$$\left|\alpha - \frac{p_m}{q_m}\right| > \frac{\gamma}{q_m^n}.$$

Отсюда следовало бы, что для *всех* m

$$\frac{\gamma}{q_m^n} < \frac{2}{q_m^{m+1}},$$

а это невозможно, если m достаточно велико. Итак, рассматриваемое число α трансцендентно.

В связи с приближением иррациональных чисел рациональными стоит упомянуть также следующую важную теорему: если α иррационально, то существует *бесконечно много* рациональных чисел p/q (p и q взаимно просты), таких, что

$$\left|\alpha - \frac{p}{q}\right| \leqslant \frac{1}{q^2}.$$

Доказательство этой теоремы интересно постольку, поскольку в нем используется широко применяемый «принцип клеток» Дирихле; он часто формулируется как утверждение о том, что если m объектов распределены по n клеткам и $m > n$, то по крайней мере одна клетка будет содержать не менее двух объектов.

Наиболее известное применение этого принципа — доказательство того, что в любом достаточно населенном городе по крайней мере два человека имеют одинаковое количество волос на голове. Для этого достаточно знать, что число волос на любой голове меньше числа жителей рассматриваемого города. Это условие выполняется, например, для Нью-Йорка, население которого составляет, грубо говоря, 8 000 000 (человек был бы раздавлен весом такого количества волос). Если каждому жителю Нью-Йорка выдать ярлычок с указанием числа волос на его голове, то по крайней мере двое получат ярлычки с одним и тем же числом, т. е. количество волос на головах этих людей одинаково.

А вот пример немного посложнее: в городе с населением не менее 21 000 жителей по крайней мере двое имеют одинаковые инициалы. (Мы предполагаем, что инициалы состоят из двух или трех букв.) В самом деле, поскольку в английском алфавите 26 букв, существует 26×26 различных двухбуквенных инициалов (A. A., A. B., ..., Y. Z., Z. Z.) и $26 \times 26 \times 26$ трехбуквенных (A. A. A., A. A. B., ..., Y. Z. Z., Z. Z. Z.). Таким образом, общее число различных инициалов равно

$$26^2 + 26^3 < 21\,000.$$

Представим себе теперь, что все жители нашего города расселены в $26^2 + 26^3$ квартир в соответствии с их инициалами. Тогда хотя бы двое из них окажутся в одной квартире и, следовательно, будут иметь одинаковые инициалы.

Вернемся к нашей теореме. Пусть Q — некоторое натуральное число. Рассмотрим числа 0, (α), (2α), ..., $(Q\alpha)$, где (α) обозначает дробную часть α; например, $(5/3) = 2/3$, $(3) = 0$, $(\sqrt{5}) = \sqrt{5} - 2$ и т. д. Рассмотрим, далее, Q интервалов («клеток»)

$$0 \leqslant x < \frac{1}{Q}, \quad \frac{1}{Q} \leqslant x < \frac{2}{Q}, \quad \ldots, \quad \frac{Q-1}{Q} \leqslant x < 1,$$

в которые должны попасть перечисленные выше $Q+1$ дробных частей $Q+1$ чисел $0, \alpha, 2\alpha, \ldots, Q\alpha$. В силу принципа Дирихле, найдется хотя бы один интервал, содержащий не менее двух дробных частей рассматриваемых чисел. Иначе говоря, найдутся два различных натуральных числа q_1 и q_2, не превосходящих Q, и натуральное s, такие, что

$$\frac{s}{Q} \leqslant (q_1\alpha) < \frac{s+1}{Q} \quad \text{и} \quad \frac{s}{Q} \leqslant (q_2\alpha) < \frac{s+1}{Q}.$$

Можно считать, что $q_2 > q_1$, и положить $q = q_2 - q_1$, так что $0 < q \leqslant Q$. Отсюда сразу следует, что $q\alpha$ есть целое число плюс или минус положительная дробь, не превосходящая $1/Q$; иными словами, существует целое p, такое, что

$$|q\alpha - p| < 1/Q.$$

Следовательно (поскольку $q \leqslant Q$),

$$\left|\alpha - \frac{p}{q}\right| < \frac{1}{Qq} \leqslant \frac{1}{q^2}.$$

Если лишь конечное число дробей $p_1/q_1, p_2/q_2, \ldots$ $\ldots, p_r/q_r$ удовлетворяет условию

$$\left|\alpha - \frac{p_i}{q_i}\right| \leqslant \frac{1}{q_i^2}, \quad i = 1, 2, \ldots, r,$$

то, поскольку α иррационально, существует целое Q, такое, что

$$\left|\alpha - \frac{p_i}{q_i}\right| > \frac{1}{Q}.$$

Повторяя предыдущие рассуждения, мы могли бы найти несократимую дробь p/q, такую, что

$$\left|\alpha - \frac{p}{q}\right| < \frac{1}{Qq} \leqslant \frac{1}{Q}.$$

Эта дробь не совпадает ни с одной из дробей p_i/q_i, $i = 1, 2, \ldots, r$, однако удовлетворяет условию

$$\left|\alpha - \frac{p}{q}\right| < \frac{1}{Qq} \leqslant \frac{1}{q^2},$$

вопреки нашему предположению, что дробями p_i/q_i, $i = 1, 2, \ldots, r$, исчерпываются *все* рациональные числа, удовлетворяющие этому условию.

Эти применения принципа Дирихле хорошо иллюстрируют природу математического творчества и математической изобретательности. Сам принцип, возможно, когда-нибудь выведет и вычислительная машина, исходя из аксиом арифметики. Однако сообразит ли машина, какое отношение имеет этот принцип, скажем, к задаче о жителях с одними и теми же инициалами? Если бы это было так, замена людей машинами стала бы, чего доброго, осуществимой!

§ 4. Трансцендентные числа: канторовское доказательство

Точную математическую формулировку понятию бесконечности дал Георг Кантор, работы которого были столь удивительны, что математики некоторое время считали их неприемлемыми. Используя его идеи, можно, например, доказать существование трансцендентных чисел, *не указывая ни одного из них.* В этом параграфе мы приведем доказательство Кантора. Главным понятием здесь является понятие *счетного множества.* Элементы такого множества можно занумеровать натуральными числами 1, 2, 3, ..., т. е. их можно «перечислить». На первый взгляд может показаться, что любое множество счетно; однако, как мы увидим ниже, в действительности дело обстоит иначе.

Рассуждения, приведенные в этом параграфе, резко отличаются от рассуждений § 3. Сравнение содержания этих двух параграфов позволяет уловить разницу между *экзистенциальным* и *конструктивным* доказательствами. На этой почве в математике произошел раскол и выделились различные направления; подробнее об этом будет сказано в гл. 2.

Как мы уже упоминали, Кантор доказал существование трансцендентных чисел, показав, что множество всех алгебраических чисел *меньше* множества всех действительных чисел. В этом доказательстве сравниваются *бесконечные множества.* Метод Кантора оказался на редкость стимулирующим и плодотворным, поэтому мы кратко его опишем.

Чтобы определить в точных терминах, что понимается под утверждением: *два множества имеют одинаковое количество элементов*, необходимо ввести понятие *взаимно однозначного соответствия между множествами.* Такое соответствие есть просто «спаривание» элементов одного множества с элементами другого, т. е. способ сопоставления каждому элементу некоторого множества *одного и только одного* элемента другого множества. Если между двумя множествами можно установить такое соответствие, то говорят, что эти множества имеют одинаковое число элементов, или *одинаковую мощность* [1]).

Бесконечное множество называется *счетным*, если оно имеет ту же мощность, что и множество натуральных чисел 1, 2, 3, ... Иными словами, множество счетно, если его элементы можно расположить в последовательность: 1-й элемент, 2-й элемент, 3-й элемент и т. д.

Можно доказать, что объединение счетного числа конечных или счетных множеств не более чем счетно. (Оно может оказаться конечным, если некоторые множества в этом объединении пусты.) Отсюда следует, что множество алгебраических уравнений любой *заданной* степени счетно (напомним, что коэффициенты такого уравнения — целые числа). Поэтому множество алгебраических уравнений *всех* степеней счетно, и значит, счетно также множество всех алгебраических чисел, являющихся корнями этих уравнений.

Допустим теперь, что множество всех действительных чисел счетно. Поскольку каждое такое число можно *единственным образом* представить в виде бесконечной десятичной дроби (например, $1{,}1 = 1{,}0999\ldots$), предположим, что *все* действительные числа записаны в последовательность

$$\begin{array}{l} c_1, c_{11}c_{12}c_{13}\ \ldots \\ c_2, c_{21}c_{22}c_{23}\ \ldots \\ c_3, c_{31}c_{32}c_{33}\ \ldots \\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot \end{array}$$

[1]) Следует отметить, что счет — это и есть установление взаимно однозначного соответствия между объектами, которые пересчитываются, и элементами некоторого *стандартного* множе-

Пусть d_1 — любая цифра, отличная от c_1, d_2 — любая цифра, отличная от c_{21}, d_3 — любая цифра, отличная от c_{32}, и т. д. Тогда действительное число

$$d_1,d_2d_3 \ldots$$

отличается от любого числа нашей последовательности, и мы приходим к противоречию. Следовательно, множество всех действительных чисел *нельзя* расположить в последовательность, и, значит, оно *несчетно*.

Контраст между методами Лиувилля и Кантора поразителен. Эти методы служат прекрасной иллюстрацией двух совершенно различных подходов к доказательству *существования* математических объектов. Метод Лиувилля чисто *конструктивный*, метод Кантора — чисто *экзистенциальный*. При первом подходе объект выявляется путем его построения, при втором — доказывается, что он существует, ибо является элементом некоторого непустого множества.

Введение чисто экзистенциальных доказательств, основанных на теории бесконечных множеств, оказало глубокое влияние на развитие математики. Это — по-видимому, единственное серьезное методологическое изменение со времен древней Греции, и оно породило — также, по-видимому, единственное — серьезное разделение математиков на группы, придерживающиеся различных взглядов на основания математики.

Чтобы ощутить, сколь далеко идущие последствия имел метод Кантора, рассмотрим следующий интересный парадокс. Ясно, что множество чисел, которые можно определить любым конечным числом слов, счетно. Но мы доказали, что множество всех действительных чисел несчетно. Значит, существуют действительные числа, которые нельзя описать никакой, хотя бы и очень длинной фразой.

Это неожиданное заключение взволновало и даже шокировало многих математиков. Великий Пуанкаре

ства. Так, при счете на пальцах устанавливают соответствие между множеством чьих-либо пальцев и элементами какого-то множества объектов.

счел его достаточным основанием, чтобы объявить «канторизму» войну, выступив против него со своей знаменитой фразой: «...n'envisager jamais que les objets susceptibles d'être defini dans un nombre fini des mots» («... никогда не рассматривайте никаких объектов, кроме тех, которые можно определить конечным числом слов»).

§ 5. Еще некоторые доказательства невозможности

Силу математики хорошо иллюстрируют доказательства невозможности. В этом параграфе мы приведем несколько примеров. Начнем с классической (восходящей еще ко временам Платона) задачи об *удвоении куба*: задан некоторый куб; требуется при помощи циркуля и линейки построить куб вдвое большего объема. Если ребро заданного куба равно a, то ребро искомого куба b должно удовлетворять уравнению $b^3 = 2a^3$, откуда $b = \sqrt[3]{2}a$. Таким образом, задача была бы решена, если бы можно было построить $\sqrt[3]{2}$, а это, как мы строго докажем несколько ниже, невозможно.

В заключение мы приведем несколько элементарных, но выразительных примеров математической невозможности в некоторых геометрических задачах.

В доказательствах невозможности наилучшим образом проявляется своеобразие и уникальный характер математического рассуждения. Утверждение, что удвоение куба (т. е. построение числа $\sqrt[3]{2}$ циркулем и линейкой) невозможно, — это не просто ссылка на то, что *в настоящее время* человек не способен справиться с такой задачей. Нет, сделанное утверждение гораздо категоричнее: оно означает, что *никогда*, ни при каких обстоятельствах, никому не удастся построить $\sqrt[3]{2}$ или разделить на три части угол общего вида, если в его распоряжении не будет других инструментов, кроме циркуля и (односторонней) линейки.

Никакая другая наука и вообще никакая другая область человеческих знаний не может даже надеяться на какие-либо столь же окончательные результаты. Не удивительно, что и сегодня некоторые люди не способны или не хотят понять, о чем идет речь, и упрямо пытаются удвоить куб, ищут трисекцию угла или квадратуру круга.

Попробуем объяснить более подробно (хотя и не исчерпывающе), в чем же дело. Пусть на плоскости задан отрезок произвольной длины; примем длину этого отрезка за единицу. Существуют стандартные приемы построения циркулем и линейкой отрезков рациональной длины p/q (p и q — взаимно простые целые числа). Назовем число α *построимым*, если циркулем и линейкой можно построить отрезок длины α. Вводя направленные отрезки, легко распространить понятие построимости и на отрицательные числа.

Если α и β построимы, то легко видеть, что $\alpha + \beta$, $\alpha - \beta$, $\alpha\beta$ и α/β тоже обладают этим свойством. Кроме того, — это иллюстрирует рис. 1 — число $\sqrt{\alpha\beta}$ тоже

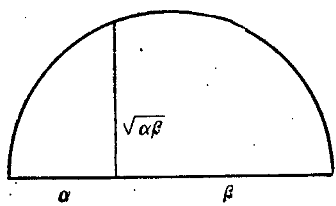


Рис. 1. Исходной фигурой для построения числа $\sqrt{\alpha\beta}$ служит полуокружность диаметра $\alpha+\beta$.

построимо. Множество построимых чисел определяется двумя следующими свойствами: 1) оно содержит все рациональные числа; 2) если ему принадлежат числа α и β, то ему принадлежат и числа $\alpha \pm \beta$, $\alpha\beta$, α/β и $\sqrt{\alpha\beta}$ (доказательство этого мы здесь опустим). Таким образом, предложение о невозможности удвоения куба сводится к утверждению о том, что $\sqrt[3]{2}$ не принадлежит множеству построимых чисел.

Общее доказательство уже содержится в зародыше в доказательстве следующего значительно более слабого утверждения: $\sqrt[3]{2}$ нельзя представить в виде дроби

$$(a+b\sqrt{c})/(d+e\sqrt{f}),$$

где a, b, c, d, e, f — рациональные числа.

Мы можем считать, что $\sqrt{c}$ и $\sqrt{f}$ *рационально независимы* (т. е. что $\sqrt{f}$ нельзя представить в виде

$$(\alpha+\beta\sqrt{c})/(\gamma+\delta\sqrt{c})$$

с рациональными α, β, γ, δ), ибо иначе можно было бы упростить наше выражение так, чтобы оно содержало только $\sqrt{c}$.

Допустим теперь, что

$$\sqrt[3]{2}=\frac{a+b\sqrt{c}}{d+e\sqrt{f}}.$$

Избавляясь от иррациональности в знаменателе (т. е. умножая числитель и знаменатель на $d-e\sqrt{f}$), мы получаем:

$$\sqrt[3]{2}=A+B\sqrt{c}+C\sqrt{f}+D\sqrt{cf},$$

где A, B, C, D снова рациональны (например, $A=ad/(d^2-e^2f)$). Заметим, что число $\sqrt{cf}$ иррационально, так как иначе $\sqrt{c}$ и $\sqrt{f}$ не были бы рационально независимыми. Теперь напишем

$$\sqrt[3]{2}=M+N\sqrt{f} \qquad (N=C+D\sqrt{c}, \quad M=A+B\sqrt{c})$$

и возведем обе части этого равенства в куб:

$$2=M^3+3MN^2f+(3M^2N+N^3f)\sqrt{f}.$$

Мы утверждаем, что коэффициент при $\sqrt{f}$ в этом выражении равен нулю, т. е.

$$3M^2N+N^3f=0.$$

В самом деле, в противном случае можно было бы разрешить последнее уравнение относительно $\sqrt{f}$:

$$\sqrt{f}=\frac{2-M^3-3MN^2f}{3M^2N+N^3f},$$

т. е. $\sqrt{f}$ выражалось бы рационально через $\sqrt{c}$, что противоречит нашему предположению о рациональной независимости чисел $\sqrt{c}$ и $\sqrt{f}$.

Итак, $3M^2N + N^3f = 0$; следовательно,

$$(M - N\sqrt{f})^3 = M^3 + 3MN^2f - (3M^2N + N^3f)\sqrt{f} = 2,$$

поэтому $M + N\sqrt{f}$ и $M - N\sqrt{f}$ — корни кубического уравнения $x^3 - 2 = 0$. Поскольку сумма (трех) корней этого уравнения равна 0, то $-2M$ тоже его корень, т. е.

$$(A + B\sqrt{c})^3 = -{}^1/_4.$$

В точности так же мы приходим к выводу, что $A + B\sqrt{c}$ и $A - B\sqrt{c}$ — корни уравнения $x^3 + {}^1/_4 = 0$, а потому и $-2A$ его корень. Следовательно, $A^3 = {}^1/_{32}$, откуда $A = \sqrt[3]{2}/4$, т. е. $\sqrt[3]{2}$ — рациональное число. Но это неверно: иррациональность $\sqrt[3]{2}$ можно доказать точно так же, как иррациональность $\sqrt{2}$.

Итак, предположив, что $\sqrt[3]{2}$ представляется в виде $\dfrac{a + b\sqrt{c}}{d + e\sqrt{f}}$ с рациональными a, b, c, d, e, f, мы пришли к противоречию; следовательно, $\sqrt[3]{2}$ нельзя представить в таком виде.

Поясним теперь коротко, как можно дополнить приведенные рассуждения, с тем чтобы доказать непостроимость $\sqrt[3]{2}$. Для начала заметим, что всякое построимое число имеет вид

$$\frac{a_0 + a_1\sqrt{P_1} + \ldots + a_k\sqrt{P_k}}{b_0 + b_1\sqrt{Q_1} + \ldots + b_l\sqrt{Q_l}},$$

где коэффициенты a_i и b_j — рациональные числа, а каждое P_i и каждое Q_j есть линейная комбинация (с рациональными коэффициентами) некоторых радикалов.

Введем понятия *степени* и *порядка* построимого числа. Сначала определим степень радикала вида $\sqrt{P}$ (или $\sqrt{Q}$). Будем говорить, что $\sqrt{P}$ имеет степень n, если P есть линейная комбинация (с рациональными коэффициентами) радикалов степени $n-1$ и ниже, причем по крайней мере один из этих радикалов имеет степень $n-1$. Например,

$$\sqrt{\frac{1}{2}+2\sqrt{\frac{1}{3}+\sqrt{2+\frac{1}{5}\sqrt{2+\sqrt{3}}}}}$$

имеет степень 5 (степень рационального числа равна 0).

Степень построимого числа равна наибольшей из степеней радикалов $\sqrt{P_1}, \ldots, \sqrt{P_k}, \sqrt{Q_1}, \ldots, \sqrt{Q_l}$. При этом подразумевается, что при вычислении степени $\sqrt{P}$ (или $\sqrt{Q}$) все радикалы приведены к простейшему виду. Так, степень радикала

$$\sqrt{1+\sqrt{2+\sqrt{9}}}$$

равна не 3, как может показаться на первый взгляд, а 2, поскольку $\sqrt{9}=3$ и

$$\sqrt{1+\sqrt{2+\sqrt{9}}}=\sqrt{1+\sqrt{5}}.$$

Если степень построимого числа равна n, то его порядок r определяется как число рационально независимых радикалов степени n. [Рационально независимыми называются радикалы, которые нельзя получить из других радикалов рациональными операциями (сложением, вычитанием, умножением и делением).] Например, рассмотренное выше выражение

$$\frac{a+b\sqrt{c}}{d+e\sqrt{f}}$$

имеет степень $n=1$ и порядок $r=2$.

Если теперь предположить, что $\sqrt[3]{2}$ — построимое число степени n и порядка r, то, рассуждая почти в точности так же, как выше, мы сможем показать, что на самом деле оно имеет порядок $r-1$. Повторяя это рассуждение, мы покажем, что порядок этого числа равен 0 и, следовательно, его степень меньше n. Таким способом мы в конце концов придем к выводу, что число $\sqrt[3]{2}$ рационально, и тем самым получим противоречие.

Технически менее сложным, но по существу столь же не простым, и, вероятно, более типичным является данное Гильбертом доказательство теоремы Штейнера о невозможности нахождения центра заданного круга при помощи одной линейки.

Что такое построение, выполненное одной линейкой? Это *конечная* последовательность шагов, на каждом из которых мы либо проводим прямую, либо находим точку пересечения двух прямых или прямой и заданной окружности. Прямую можно проводить через две точки, выбранные с той или иной степенью произвола; например, один из шагов может состоять в выборе двух произвольных точек на заданной окружности и соединении их хордой или в проведении прямой через две точки, одна из которых (или обе они) получена на каком-то более раннем шаге построения как пересечение двух прямых или прямой и заданной окружности. Эта последовательность шагов должна привести в конце концов к некоторой точке, относительно которой можно доказать, что она является центром нашего круга.

Допустим, что построение выполняется на некоторой плоскости P_1, и представим себе такое преобразование, или отображение, T плоскости P_1 в другую плоскость P_2, что

(a) прямые плоскости P_1 переходят в прямые плоскости P_2, т. е. если точки p, q, r, ... лежат на одной прямой l плоскости P_1, то их «образы» $T(p)$, $T(q)$, $T(r)$, ... лежат на прямой $T(l)$ плоскости P_2;

(b) заданная окружность C плоскости P_1 переходит в некоторую окружность $T(C)$ плоскости P_2.

Каждый шаг построения, выполненного на P_1, копируется соответствующим построением на P_2, и если бы первое построение заканчивалось в центре O окружности C, то «построение-образ» должно было бы закончиться *в центре* $T(O)$ окружности $T(C)$. Поэтому если удастся найти преобразование T, удовлетворяющее условиям (a) и (b), но такое, что $T(O)$ не является центром $T(C)$, то невозможность построения центра круга одной только линейкой будет доказана.

Такое преобразование T показано на рис. 2; оно называется центральной проекцией с центром S. Центральные проекции искажают расстояния и могут переводить эллипсы в гиперболы. Однако очень многие интересные и важные свойства геометрических фигур остаются неизменными (инвариантны) при этих отображениях. Изучением таких свойств занимается проективная геометрия. Если евклидову геометрию можно представлять себе как геометрию циркуля и линейки, то проективная геометрия — это геометрия одной линейки. Таким образом, построения при помощи одной линейки должны быть *проективно инвариантны*. С другой стороны, отношение между кругом и его центром не является проективно инвариантным и, следовательно, не описывается в чисто проективных терминах.

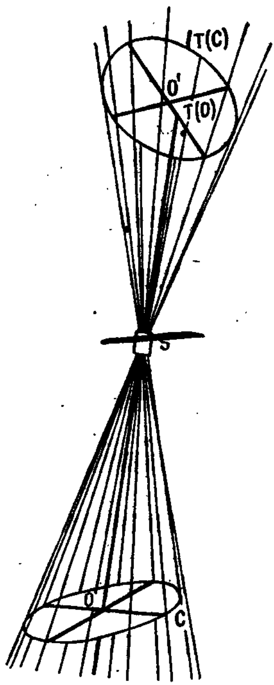


Рис. 2.

Другие примеры доказательств невозможности мы возьмем из комбинаторного анализа, демонстрируя тем самым его принадлежность к математике.

Рассмотрим квадрат, разделенный на 64 равных квадрата, т. е. шахматную доску. Отбросим верхний правый и нижний левый угловые квадраты. Можно ли покрыть остальные 62 квадрата 31 костью домино (т. е. прямоугольниками, состоящими из двух смежных квадратов)? Вот элегантное доказательство невозможности такого покрытия. Если раскрасить квадратики в черный и белый цвет, как на шахматной доске, то оба выброшенные поля будут одного цвета, скажем белого. Поскольку кость домино покрывает одно черное и одно белое поле, а на оставшейся части доски черных полей на два больше, чем белых, такое покрытие неосуществимо. Невозможность стала совершенно очевидной, как только мы додумались раскрасить квадратики, хотя исходная картина разделенного квадрата не предполагала никакой раскраски.

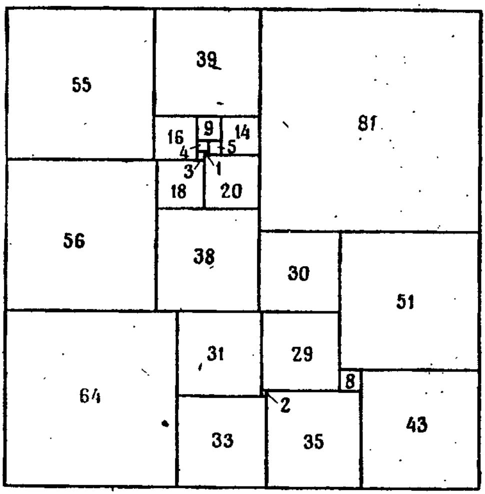


Рис. 3.

Вот еще одно доказательство невозможности, тоже связанное с задачей комбинаторной геометрии (части комбинаторного анализа).

Квадрат можно разрезать на конечное число квадратов, среди которых нет двух одинаковых[1]). На рис. 3 приводится простое доказательство этого факта.

---

[1]) Существует интересная теория «квадрирования квадрата», т. е. нахождения всех возможных разбиений квадрата на *неравные* квадраты с параллельными сторонами. Она тесно связана с теорией Кирхгофа о токах в электрических цепях. Это еще одна иллюстрация замечательных и совершенно неожиданных связей, которыми полна математика.

Возникает вопрос, можно ли получить аналогичное разбиение куба, т. е. можно ли разрезать куб на конечное число неравных кубов? Ответ на этот вопрос отрицателен. Допустим, что такое разбиение существует. Тогда нижнее основание исходного куба будет разбито нижними гранями этих неравных кубов на неравные квадраты. Рассмотрим наименьший из этих квадратов. Он не может лежать в углу и не может примыкать к стороне большого квадрата, ибо в этом случае два (или один) соседних квадрата высовывались бы из-за него так, что к четвертой его стороне нельзя было бы приставить квадрат с большей стороной. Итак, самый маленький квадрат должен находиться где-то в середине нижней грани исходного куба. Рассмотрим теперь кубы, расположенные по четыре стороны от этого квадратика. Их стороны больше его стороны, следовательно, все эти кубы возвышаются над верхней гранью куба, нижней гранью которого служит наш маленький квадратик. Поэтому на этой верхней грани могут стоять лишь кубы меньшего размера. Рассмотрим наименьший из покрывающих ее квадратов и начнем все сначала. Мы придем к выводу, что этот процесс не может завершиться после конечного числа шагов, так как, повторяя наши рассуждения, мы будем последовательно получать все меньшие и меньшие квадраты на все более высоких уровнях. Это рассуждение и доказывает невозможность разбиения куба на конечное число меньших и не равных друг другу кубов.

§ 6. Лемма Шпернера

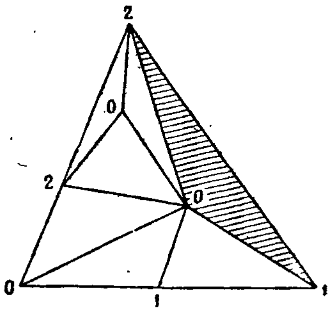
Математические доказательства невозможности носят характер утверждения «наверняка нет»; родственные им доказательства существования утверждают «наверняка да». Яркий пример математического доказательства существования дает лемма Шпернера. Она относится к важному разделу математики — *комбинаторной топологии*, в которой геометрические объекты классифицируются по свойствам, сохраняющимся при растяжениях и произвольных непрерывных (или гладких) деформациях. Например, в топологии не различают (относят к одному и тому же классу геометрических образов) круг и квадрат; это связано с тем, что можно найти взаимно однозначное соответствие между точками круга и квадрата, или *отображение*, переводящее квадрат в круг, причем такое соответствие, или отображение, непрерывно (т. е. близкие точки соответствуют близким). Эти отображения обладают одним замечательным свойством, которое выражает так называемая теорема Брауэра о неподвижной точке — непосредственное следствие леммы Шпернера.

Лемма Шпернера, одно из самых мощных средств комбинаторной топологии, касается разбиений треугольника на меньшие треугольники и нумерации вершин треугольников разбиения.

Представим себе треугольник или один из его многомерных аналогов, например тетраэдр в трехмерном пространстве. Допустим, что этот треугольник (или многомерный симплекс) «симплициально» разбит на конечное число меньших треугольников (симплексов), т. е. разбиение таково, что меньшие треугольники могут иметь общими либо целую сторону, либо только вершину. Предположим далее, что мы занумеровали вершины исходного треугольника числами 0, 1, 2 (см. рис. 4). Вершины разбиения, лежащие на сторонах исходного треугольника, нумеруются так: точка на стороне 0, 1 получает номер 0 или 1, на стороне 0, 2 — номер 0 или 2, на стороне 1, 2 — номер 1 или 2; в остальном нумерация произвольна. Тогда справедлива лемма Шпернера: *вершины одного из меньших треугольников (симплексов) занумерованы всеми числами* 0, 1, 2. Оказывается, что число таких выделенных треугольников (симплексов) разбиения нечетно!

Аналогичная теорема имеет место для симплициальных разбиений тетраэдра, вершины которого нумеруются числами 0, 1, 2, 3.

Докажем сначала эту лемму для симплициального разбиения отрезка, т. е. для разбиения отрезка на меньшие отрезки. Концы исходного отрезка нумеруются числами 0, 1; каждый конец отрезков разбиения — либо числом 0, либо числом 1. Для каждого отрезка разбиения α обозначим через $\nu(\alpha)$ число его концов, занумерованных 0; таким образом, $\nu(\alpha)=0$, если оба конца имеют номер 1, $\nu(\alpha)=1$, если один конец имеет номер 0, а другой — номер 1, и $\nu(\alpha)=2$, если оба конца имеют номер 0. Отрезок α называется



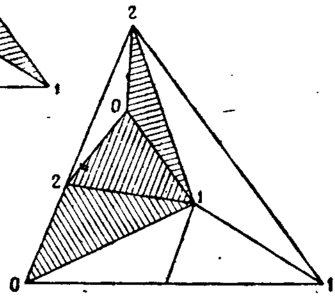


Рис. 4. Два варианта выбора обозначений вершин для одного и того же симплициального разбиения, удовлетворяющих условиям леммы Шпернера. Заштрихованные треугольники обозначены всеми числами 0, 1, 2.

выделенным, если один из его концов имеет номер 0, а другой 1, т. е. если $\nu(\alpha)=1$; при этом для каждого невыделенного отрезка $\nu(\alpha)$ четно ($\nu(\alpha)=0$ или 2). Пусть m — число выделенных отрезков. Тогда

$$m-\sum\nu(\alpha)=\text{четное число},$$

где $\sum$ означает суммирование по всем отрезкам разбиения.

Заметим, что всякий внутренний конец, отмеченный цифрой 0, засчитывается в $\sum\nu(\alpha)$ дважды (ибо он является концом в точности двух отрезков) и что лишь один конец исходного отрезка имеет номер 0. Значит, $\sum\nu(\alpha)$ равна 1 плюс четное число, т. е. нечетна. Поскольку m отличается от $\sum \nu(\alpha)$ на четное число, m нечетно, и тем самым доказательство леммы Шпернера в простейшем (одномерном) случае завершено.

В случае треугольника поступим аналогично и для каждого треугольника разбиения α обозначим через $\nu(\alpha)$ число его сторон с концами (вершинами) 0 и 1. Треугольник α назовем выделенным, если его вершины отмечены всеми допустимыми цифрами, т. е. цифрами 0, 1, 2; при этом важно заметить, что $\nu(\alpha)$ четно для всех α, за исключением выделенных, и $\nu(\alpha)=1$ для всех выделенных α. Обозначим снова через m число выделенных подсимплексов. Тогда

$$m-\sum \nu(\alpha)=\text{четное число.}$$

Но каждая внутренняя сторона (0, 1) считается дважды, поэтому

$$\sum \nu(\alpha)=\text{четное число}+\text{число сторон (0, 1), лежащих на стороне (0, 1) исходного треугольника.}$$

Но последнее слагаемое нечетно, так как здесь мы имеем дело с симплициальным разбиением отрезка, для которого лемма Шпернера уже доказана. Итак, m нечетно.

Заметим, что доказательство для треугольника сводится к доказательству для отрезка, т. е. двумерный вариант леммы Шпернера сводится к одномерному. Аналогично трехмерный случай (симплициальное разбиение тетраэдра) можно свести к двумерному, или одномерному, и этот процесс можно продолжать бесконечно, если знать, как определять многомерные симплексы.

Этот весьма частный чисто комбинаторный результат имеет важные и неожиданные следствия. В качестве иллюстрации его применения опишем вкратце доказательство известной теоремы Брауэра.

В одномерном случае эта теорема утверждает, что если некоторое отображение T непрерывно переводит отрезок в себя, т. е. каждой точке p отрезка ставит в соответствие точку $T(p)$ того же отрезка, причем

$T(q)$ можно сделать сколь угодно близкой к $T(p)$, взяв q достаточно близкой к p, то существует по крайней мере одна точка p_0, остающаяся неподвижной:

$$T(p_0) = p_0.$$

Это доказывается так. Рассмотрим симплициальное разбиение данного отрезка, т. е. попросту возьмем некоторое количество его точек. Разобьем эти точки на два множества. Пусть первое состоит из тех точек, расстояние которых от левого конца отрезка (обозначенного цифрой 0) не уменьшилось после отображения. Ясно, что точка 0 входит в это множество; обозначим все остальные его точки также цифрой 0. Во второе множество отнесем точки, расстояние которых от правого конца не уменьшилось; обозначим их 1. Тогда по лемме Шпернера существует отрезок разбиения с концами 0 и 1. Это означает, что существуют две близкие точки, одна из которых в результате нашего отображения не приблизилась к левому концу, а другая — соседняя — не приблизилась к правому. Эти точки можно сделать сколь угодно близкими, так как можно выбрать произвольно мелкое разбиение. Переходя к пределу, мы заключаем, что должна существовать по крайней мере одна точка, расстояние которой от обоих концов не уменьшилось, т. е. точка, которая вообще не сдвинулась с места. В двумерном и многомерном случаях рассуждение по существу остается тем же.

Теоремы о неподвижной точке — одно из самых мощных средств современной математики. Данное выше доказательство было получено из «конечных» соображений, связанных с «теоремой существования» симплекса разбиения с полным множеством номеров вершин, т. е. *невозможностью* такой нумерации вершин симплексов разбиения, чтобы такого выделенного симплекса не существовало.

Из трех последних примеров задача о домино на шахматной доске и задача о делении куба на неодинаковые меньшие кубики могут показаться лишь любопытными головоломками (и, видимо, так оно и есть). Лемма Шпернера представляется гораздо более важным и глубоким результатом, имеющим многие приложения. И все же формально в этих примерах есть нечто общее: все они могут быть связаны со счетом числа некоторых образов.

Очень часто именно многочисленные приложения в различных и кажущихся совсем не связанными областях математики придают значительность и красоту найденному результату. По словам Декарта, «рассматривая примеры, можно создать *метод*». Однако, вообще говоря, нельзя формально провести черту между величественным и смешным. Наши критерии, касающиеся этого различия, отчасти носят эстетический характер; они определяются тем, насколько хорошо те или иные теоремы, идеи и методы «работают» в других ситуациях.

§ 7. Искусство и наука счета

В этом параграфе мы рассмотрим одну элементарную задачу на подсчет числа вариантов, которая с неизбежностью приведет нас к комплексным числам, введенным в § 2. Это может служить иллюстрацией того, сколь типично для задач подсчета привлечение самых разнообразных понятий, идей и методов.

Счет — такой простой и привычный с детства процесс, что приходится только удивляться, сколько интересных, важных и сложных задач с ним связано.

Рассмотрим задачу о том, сколькими различными способами можно разменять доллар, т. е. сколько различных решений имеет уравнение

$$100 = l_1 + 5l_2 + 10l_3 + 25l_4 + 50l_5,$$

где под «решением» мы понимаем пятерку неотрицательных целых чисел $(l_1, l_2, l_3, l_4, l_5)$. Здесь l_1 — число монет в 1 цент, l_2 — число никелей и т. д.[1]).

[1]) В США находятся в обращении следующие монеты: 1 цент, 1 никель = 5 центов, 1 дайм = 10 центов, 1 квартер = = 25 центов, 1 полудоллар = 50 центов, 1 доллар = 100 центов. — *Прим. перев.*

Попытка просто перечислить все различные возможности очень скоро обескураживает, поскольку мы убеждаемся в почти полной ее безнадежности.

Перефразируем нашу задачу. Рассмотрим ряды

$$1+x+x^2+x^3+\ldots,$$
$$1+x^5+x^{10}+x^{15}+\ldots,$$
$$1+x^{10}+x^{20}+x^{30}+\ldots,$$
$$1+x^{25}+x^{50}+x^{75}+\ldots,$$
$$1+x^{50}+x^{100}+x^{150}+\ldots,$$

где показатели степени в первом ряде — целые числа, кратные 1, во втором — числа, кратные 5, и т. д.

Перемножив эти ряды формально (т. е. пренебрегая тем, что они имеют бесконечную длину, и обращаясь с ними, как с обычными многочленами), мы получим ряд вида

$$1+A_1x+A_2x^2+A_3x^3+\ldots;$$

можно показать, что A_m есть число различных способов, которыми можно представить число m в виде $l_1+5l_2+10l_3+25l_4+50l_5$. В частности, A_{100} — искомое число различных способов образовать сумму в один доллар из более мелких монет.

Заметим, что каждый из рядов представляет собой сумму членов бесконечной геометрической прогрессии. Поэтому их произведение равно (снова формально)

$$\frac{1}{(1-x)(1-x^5)(1-x^{10})(1-x^{25})(1-x^{50})}.$$

Прервем на время наши рассуждения и рассмотрим гораздо более простую задачу. Допустим, что мы хотим найти число различных решений уравнения

$$100=l_1+2l_2.$$

В этом случае мы пришли бы к выражению

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

Попытаемся разложить его на элементарные дроби, т. е. найти такие числа a, b, c, чтобы тождественно по x (т. е. для всех без исключения x) выполнялось равенство

$$\frac{1}{(1-x)^2(1+x)} \equiv \frac{a}{(1-x)^2} + \frac{b}{1-x} + \frac{c}{1+x}.$$

Мы получим тождество

$$1 \equiv a(1+x) + b(1-x)(1+x) + c(1-x)^2,$$

которое приводит к системе трех линейных уравнений

$$-b+c=0,$$
$$a-2c=0,$$
$$a+b+c=1.$$

Следовательно, $a = {}^1/_2$, $b = {}^1/_4$, $c = {}^1/_4$, откуда

$$\frac{1}{(1-x)^2(1+x)} = \frac{1}{2}\frac{1}{(1-x)^2} + \frac{1}{4}\frac{1}{1-x} + \frac{1}{4}\frac{1}{1+x}.$$

Мы имеем

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

и, следовательно, коэффициент при x^{100} в произведении

$$(1+x+x^2+\ldots)(1+x^2+x^4+x^6+\ldots)$$

равен

$$\frac{1}{2}(101) + \frac{1}{4}(1) + \frac{1}{4}(1) = 51.$$

Для решения этой упрощенной задачи вся наша техника рядов и элементарных дробей на самом деле не нужна. Мы могли бы сразу заметить, что l_1 должно быть четным (поскольку $100 - 2l_2$ четно) и что от 0 до 100 включительно имеется ровно 51 четное число. Однако, решив простую задачу излишне сложным способом, мы оказались в значительно лучшем положении, чем раньше, ибо поняли, как следует обращаться с задачей более трудной. Заметим, что число монет в 1 цент должно делиться на 5; следовательно, число способов размена доллара A_{100} равно

$$B_{100} + B_{95} + B_{90} + \ldots + B_5 + B_0,$$

где B_m — число решений уравнения

$$m = k_1 5 + k_2 10 + k_3 25 + k_4 50,$$

или, что эквивалентно, число решений (в неотрицательных целых числах) уравнения

$$\frac{m}{5} = k_1 + 2k_2 + 5k_3 + 10k_4$$

(ведь все m делятся на 5). Отсюда следует, что A_{100} равняется сумме коэффициентов при членах

$$x^0 = 1,\ x,\ x^2,\ \ldots,\ x^{19},\ x^{20}$$

в произведении

$$(1 + x + x^2 + \ldots)(1 + x^2 + x^4 + \ldots)(1 + x^5 + x^{10} + \ldots) \times \\ \times (1 + x^{10} + x^{20} + \ldots) = \frac{1}{(1-x)(1-x^2)} \frac{1}{(1-x^5)(1-x^{10})}.$$

Как мы уже видели,

$$\frac{1}{(1-x)(1-x^2)} = \frac{1}{2}\frac{1}{(1-x)^2} + \frac{1}{4}\frac{1}{1-x} + \frac{1}{4}\frac{1}{1+x} = \\ = 1 + x + 2x^2 + 2x^3 + 3x^4 + 3x^5 + \ldots.$$

Подставляя сюда x^5 вместо x, получаем

$$\frac{1}{(1-x^5)(1-x^{10})} = 1 + x^5 + 2x^{10} + 2x^{15} + 3x^{20} + \ldots.$$

Таким образом,

$$(1 + x + x^2 + \ldots)(1 + x^2 + x^4 + \ldots)(1 + x^5 + x^{10} + \ldots) \times \\ \times (1 + x^{10} + x^{20} + \ldots) = \\ = (1 + x + 2x^2 + 2x^3 + 3x^4 + 3x^5 + \ldots) \times \\ \times (1 + x^5 + 2x^{10} + 2x^{15} + 3x^{20} + \ldots).$$

Выполняя умножение и складывая коэффициенты при степенях x от нулевой до 20-й включительно, мы, изрядно потрудившись, получим в ответе 292. Итак, существует 292 различных способа разменять доллар!

Идея использовать для подсчета степенные ряды (т. е. ряды вида $a_0 + a_1x + a_2x^2 + \ldots$ — так сказать, многочлены бесконечной длины) оказалась крайне плодотворной.

Решая задачу о размене доллара, мы воспользовались специальными свойствами монетной системы США. Например, мы знали, что достоинство любой монеты, кроме цента, делится на 5, и что дайм вдвое ценнее никеля, а полудоллар — квартера. Разумное обращение с такими случайными свойствами позволило нам во всей полноте использовать делимость целых чисел, сэкономив тем самым часть труда. Однако это могло в какой-то степени скрыть полную мощность и общность метода.

Чтобы лучше понять, о чем идет речь, рассмотрим задачу отыскания числа A_n решений (в неотрицательных целых числах) уравнения

$$n = l_1 + 2l_2 + 3l_3,$$

где n — неотрицательное целое число. Как и выше, мы придем к нахождению коэффициента при x^n в разложении произведения

$$\frac{1}{(1-x)(1-x^2)(1-x^3)}.$$

Попытаемся снова перейти к элементарным дробям, представив сначала знаменатель в виде

$$(1-x)(1-x^2)(1-x^3) = (1-x)^3(1+x)(1+x+x^2).$$

Однако теперь, чтобы разложить на линейные множители трехчлен $1 + x + x^2$, нам потребуются *комплексные числа*. В самом деле,

$$x^2 + x + 1 = (\alpha x + 1)(\bar{\alpha} x + 1),$$

где $\alpha = \frac{1}{2} + \frac{\sqrt{3}}{2} i$, $\bar{\alpha} = \frac{1}{2} - \frac{\sqrt{3}}{2} i$, $i^2 = -1$.

Искомое разложение на элементарные дроби имеет вид

$$\frac{1}{(1-x)(1-x^2)(1-x^3)} \equiv \frac{a}{(1-x)^3} + \frac{b}{(1-x)^2} + \frac{c}{1-x} + \frac{d}{1+x} + \frac{e}{1+\alpha x} + \frac{f}{1+\bar{\alpha} x};$$

a, b, c, d, e, f можно найти, решив систему шести линейных уравнений. Убедившись в том, что

$$\frac{1}{(1-x)^3} = \frac{1}{2}(1 \cdot 2 + 2 \cdot 3x + 3 \cdot 4x^2 + 4 \cdot 5x^3 + \ldots),$$

приняв во внимание уже упоминавшуюся формулу

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

и применив несколько раз формулу суммы геометрической прогрессии

$$\frac{1}{1-q} = 1 + q + q^2 + \ldots$$

(для $q = x$, $q = -x$, $q = -\alpha x$, $q = -\bar{\alpha} x$), получим

$$A_n = \frac{a}{2}(n+1)(n+2) + b(n+1) + c + d(-1)^n + e(-1)^n \alpha^n + f(-1)^n \bar{\alpha}^n.$$

Ответ остался незавершенным, потому что не найдены коэффициенты $a, b, \ldots, f$; их, однако, можно определить вполне стандартными приемами, и этот подсчет не имеет никакого отношения к нашей главной теме.

Самое поразительное в таком решении — это появление комплексных чисел в задаче, связанной лишь с подсчетом (и, следовательно, лишь с целыми числами). По этой причине стоит, по-видимому, в порядке отступления поговорить немного о природе и эволюции чисел.

§ 8. Отступление о числовых системах и о функциях

Разовьем здесь подробнее схему разных числовых систем, набросок которой мы привели в начале § 2.

Следуя Кронекеру, будем рассматривать натуральные числа как «данные богом» (хотя их можно было бы построить, исходя из более простых теоретико-множественных и логических понятий). Мы заметим, что сложение и умножение натуральных чисел приводят снова к натуральным числам, и поэтому до тех пор, пока рассматриваются только эти две операции, натуральные числа образуют замкнутый в себе мир — мир натуральных чисел.

Однако уже вычитание вынуждает нас покинуть этот уютный замкнутый мир и выйти за его пределы. В самом деле, даже такое простое уравнение, как

$$3+x=2,$$

не имеет решения в области натуральных чисел.

Мы вводим 0 и отрицательные числа как раз для того, чтобы сделать вычитание всегда возможным. Расширяя числовую систему, мы расширяем и область применимости операций. При этом нам приходится заботиться о том, чтобы сохранились их обычные свойства (ассоциативность, коммутативность, дистрибутивность). Именно поэтому произведение любого числа на 0 всегда считается равным нулю, а произведение двух отрицательных чисел — положительным[1]).

Расширив числовую систему так, чтобы она включала 0 и отрицательные числа, мы обнаружим, что в полученной таким образом области целых чисел уравнение вида

$$ax=b, \quad a\neq 0,$$

(где a и b — целые) может оказаться неразрешимым. Поэтому мы снова расширяем эту систему, присоединяя к ней дроби, числители и знаменатели которых — целые числа (положительные или отрицательные). После этого обычным образом определяются операции сложения и умножения, причем снова проявляется забота о том, чтобы сохранялись их основные

---

[1]) Пусть a — некоторое целое число. Тогда $0a=(1-1)a=a-a=0$, если потребовать (а мы так и делаем) дистрибутивность умножения относительно вычитания. Аналогично, $2(1-1)=0$ и, следовательно, $(-1)2=-2$; далее, $(2+(-2))(-1)=0$, откуда $-2+(-2)(-1)=0$ и $(-2)(-1)=2$.

свойства (ассоциативность, коммутативность и дистрибутивность). Теперь всегда возможны вычитание и деление (кроме деления на 0), и мы наконец-то получаем множество, *замкнутое* относительно *всех четырех* арифметических операций, — множество рациональных чисел. К сожалению (а может быть, к счастью!), нам не хватает и рациональных чисел. Простая задача — найти отношение длины гипотенузы равнобедренного прямоугольного треугольника к длине катета — приводит к ответу $\sqrt{2}$, а, как мы уже знаем, это число не рационально.

Расширить числовую систему, дополнив ее иррациональными числами (типа $\sqrt{2}$) было задачей совсем другого рода, чем все предшествующие, несравненно более тонкой и сложной, ибо для этого потребовалось оперировать с бесконечно малыми величинами.

Грубо говоря, можно поступать так. Рациональные числа записываются в виде десятичных дробей, которые либо кончаются какой-то цифрой (например, 2,13), либо имеют вид

$$a_0,a_1a_2\ldots a_mb_1b_2\ldots b_nb_1b_2\ldots b_n\ldots,$$

где, начиная с некоторого места, все время повторяется один и тот же набор цифр. Например,

$$\frac{1}{7}=0,142857\,142857\,142857\ldots.$$

Иррациональные числа можно определить как бесконечные десятичные дроби, не имеющие такого постоянно повторяющегося набора цифр.

Но бесконечная десятичная дробь $a_0,a_1a_2a_3\ldots$ — это всего лишь другой способ записи *бесконечного ряда*

$$a_0+\frac{a_1}{10}+\frac{a_2}{10^2}+\frac{a_3}{10^3}+\ldots,$$

а с такими рядами обращались весьма остроумно еще со времен парадокса Зенона об Ахиллесе и черепахе.

Поскольку $0 \leqslant a_m \leqslant 9$ для любого m, при $k=2, 3, \ldots$ мы имеем

$$\frac{a_{n+1}}{10^{n+1}}+\frac{a_{n+2}}{10^{n+2}}+\ldots+\frac{a_{n+k}}{10^{n+k}} \leqslant$$
$$\leqslant \frac{9}{10^n}\left(\frac{1}{10}+\frac{1}{10^2}+\ldots\right) \leqslant \frac{9}{10^n}\cdot\frac{1}{9}=\frac{1}{10^n},$$

и, следовательно, рассматриваемый бесконечный ряд определяет последовательность вложенных отрезков I_n с концами

$$a_0+\frac{a_1}{10}+\ldots+\frac{a_n}{10^n} \quad \text{и} \quad a_0+\frac{a_1}{10}+\ldots+\frac{a_n}{10^n}+\frac{1}{10^n}.$$

Будем считать, что иррациональное число *определено*, если оно с любой точностью может быть приближено конечными десятичными дробями; тогда написанный выше бесконечный ряд, или, что эквивалентно, соответствующая ему последовательность вложенных отрезков, действительно определяет некоторое число.

Далее снова вводятся арифметические операции (со всеми указанными выше предосторожностями), и множество действительных (т. е. рациональных и иррациональных) чисел становится замкнутым относительно всех четырех операций (деление на 0, конечно, запрещается). С другой стороны, в отличие от рациональных, действительные числа обладают так называемым «дедекиндовым свойством»: если разбить их на два *непустых* класса A и B так, чтобы любое действительное число попало либо в A, либо в B и чтобы каждое число из A было меньше каждого числа из B, то либо в A имеется наибольшее число, либо в B — наименьшее. Дедекиндово свойство выражает тот факт, что действительные числа образуют *континуум*.

Хотя строгая теория действительных чисел была построена лишь в самом конце 19-го века (в основном Дедекиндом и Кантором; наш набросок по духу ближе к канторовской теории), математики задолго до этого не испытывали никаких сомнений, свободно (и, возможно, несколько «некритично») пользуясь действительными числами.

Однако процесс расширения числовой системы не был завершен. Простые квадратные уравнения не имеют действительных корней. Например, для уравнения $x^2 + 2x + 2 = 0$ стандартная формула дает ответ

$$x_1 = 1 - \sqrt{-1} \quad \text{и} \quad x_2 = 1 + \sqrt{-1}.$$

Чтобы действовать в такой ситуации, были введены символы вида $a + bi$ с действительными a и b. Сложение и вычитание таких символов определяются очевидным образом, а при умножении двух таких «чисел» следуют обычным правилам алгебры и лишь в самом конце заменяют i^2 на -1. Так, например,

$$(2 - i)(3 + 2i) = 6 - 3i + 4i - 2i^2 = 6 + i - 2(-1) = 8 + i.$$

Довольно любопытно, что комплексные числа (т. е. символы вида $a + bi$ со сложением и умножением, определенными так, как сказано выше) долго причиняли математикам некоторые неудобства, хотя к началу 19-го века эти числа были уже полностью освоены математикой. Определенные сомнения и опасения остаются, должно быть, и сейчас, ибо до самого недавнего времени учебные программы средней школы в основном обходились без комплексных чисел.

Возможно, первоначальное название этих чисел — мнимые — способствовало созданию вокруг них легкого ореола таинственности; с другой стороны, видимо, трудно было расстаться с привычным представлением о том, что числа должны выражать результаты каких-то измерений, комплексные же числа получили подобное истолкование сравнительно поздно.

Как бы то ни было, расширение числовой системы присоединением к ней комплексных чисел принесло математике неисчислимую выгоду.

Самым большим чудом было, вероятно, то, что комплексные числа, введенные только для того, чтобы стали разрешимы все квадратные уравнения с действительными коэффициентами, сделали разрешимыми все вообще *алгебраические уравнения* (даже имеющие комплексные коэффициенты).

Этот замечательный факт, известный как *основная теорема алгебры*, можно сформулировать следующим образом: любое уравнение n-й степени

$$a_n z^n + a_{n-1} z^{n-1} + \ldots + a_0 = 0, \quad a_n \neq 0,$$

с произвольными комплексными коэффициентами имеет n комплексных корней (не обязательно различных). Другими словами, существуют n комплексных чисел $z_1, z_2, \ldots, z_n$, таких, что

$$a_n z^n + a_{n-1} z^{n-1} + \ldots + a_0 \equiv \\ \equiv a_n (z - z_1)(z - z_2) \ldots (z - z_n).$$

Таким образом, любой многочлен с комплексными коэффициентами можно разложить на линейные множители; именно знание этого факта позволило нам воспользоваться разложением на элементарные дроби в рассмотренной в § 7 задаче подсчета. Если и удивляться тому, что в формулу для A_n вошли комплексные числа, то вовсе не из-за смутного ощущения сверхъестественности помощи, которую оказали нам «мнимые величины» при решении конкретной реальной задачи. В математике нет места подобному простодушному мистицизму. Поистине удивительно то, что понятие, введенное для того, чтобы сделать разрешимыми все квадратные уравнения, неожиданно сыграло решающую роль в совсем другой ситуации.

Комплексные числа удобно интерпретировать как точки на плоскости. В самом деле, если раз и навсегда выбраны координатные оси x и y, то комплексное число $z = x + iy$ представляется точкой (x, y).

Введение комплексных чисел положило начало изучению *функций комплексного переменного*, т. е. отображений множеств комплексных чисел в множество комплексных чисел.

Простейшие функции комплексного переменного обладают многими поразительными свойствами. Например, функция

$$w = \frac{1+z}{1-z}$$

переводит внутренность круга радиуса 1 с центром в начале координат ($z=0$) в полуплоскость справа от оси y (см. рис. 5).

Поскольку умножение на i геометрически соответствует повороту плоскости на 90° против часовой стрелки, функция

$$w = i\frac{1+z}{1-z}$$

переводит единичный круг с центром в начале координат в верхнюю полуплоскость.

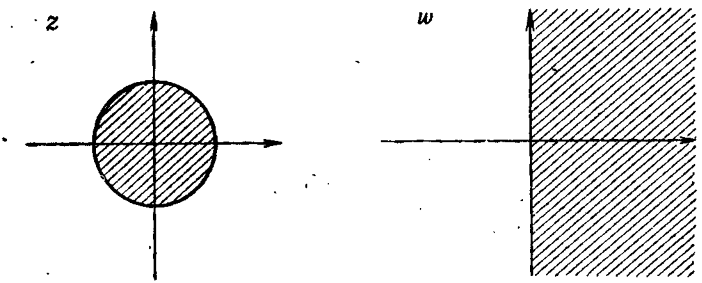


Рис. 5. Функция $w=\frac{1+z}{1-z}$ отображает заштрихованную область плоскости z — единичный круг — в заштрихованную правую полуплоскость плоскости w.

Как только стали рассматривать функции комплексного переменного, возникла идея развить для них дифференциальное и интегральное исчисления. Особый интерес представляют функции, имеющие производную, т. е. такие, для которых при каждом z из некоторой области существует предел

$$\lim_{\Delta z\to 0}\frac{f(z+\Delta z)-f(z)}{\Delta z}.$$

Требование, чтобы этот предел не зависел от способа стремления к нулю комплексного приращения Δz, налагает настолько сильные ограничения на функцию f, что если она имеет в некоторой области производную первого порядка, то в этой области существуют ее производные *всех порядков.* Более того, в окрестности каждой точки z_0 из этой области функция f представляется в виде степенного ряда, т. е.

$$f(z) = a_0 + a_1(z - z_0) + a_2(z - z_0)^2 + \ldots,$$

и этот ряд сходится в некотором круге с центром z_0. (Радиус такого «круга сходимости» может оказаться бесконечным, и тогда f называется *целой функцией.*) Функции, имеющие производную в некоторой области, называются *аналитическими* в этой области. Теория аналитических функций — одна из важнейших и красивейших глав математики. Аналитические функции проникли почти в каждый закоулок современной математики и физики, — от гидродинамики до теории чисел и от квантовой механики до топологии.

Эволюция комплексного анализа от его скромного начала — решения квадратных уравнений — до величественного здания теории аналитических функций — еще одно подтверждение жизнеспособности математических понятий.

§ 9. Искусство и наука счета (продолжение)

Мы уже несколько раз видели, что рядом с задачами, представляющими собой не более чем легко разрешимые головоломки, могут возникать по-настоящему трудные и глубокие проблемы.

Задачу о размене доллара или родственную ей задачу нахождения числа решений (в неотрицательных целых числах) уравнения

$$n = l_1 + 2l_2 + 3l_3$$

совсем легко решить, если догадаться воспользоваться степенными рядами (так называемыми *производящими функциями*).

Несравненно более трудной является задача о числе $p(n)$ решений уравнения

$$n = l_1 + 2l_2 + 3l_3 + \ldots,$$

т. е. о числе способов разбиения (или «размена») целого числа n на любые меньшие его целые положительные числа. Это известная задача partitio numerorum Эйлера.

При помощи производящих функций мы получаем

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

и поскольку на этот раз размеры «порций», на которые разбивается n, не ограничены, в правой части стоит бесконечное произведение. Теперь мы уже не можем просто разложить его на элементарные дроби. Лишь в 1934 г. Радемахер нашел красивую (но довольно сложную) формулу для $p(n)$, с большой изобретательностью и изяществом применив комплексный анализ. А именно, Радемахер обобщил метод, при помощи которого Харди и Рамануджан (1917) получили свою знаменитую асимптотическую формулу

$$p(n)\sim\frac{1}{4n\sqrt{3}}\,e^{\pi\sqrt{2n/3}}.$$

(Стоит заметить, что $p(n)$ очень быстро возрастает с ростом n; например, $p(200)$ — тринадцатизначное число!)

Вопрос о partitio numerorum в аддитивной теории чисел связан с известной задачей представления целых чисел в виде суммы квадратов.

В конце 18-го века Лагранж доказал, что всякое положительное число есть сумма четырех квадратов (трех квадратов недостаточно; например, 15 не является суммой трех квадратов). Эта теорема «о четырех квадратах», намного перекрытая позднее работами Якоби, до сих пор входит в число величайших достижений математики.

Якоби продвинулся гораздо дальше: он определил, сколькими способами число может быть записано в виде суммы четырех квадратов. Чтобы облегчить этот подсчет, удобно рассматривать представления, отличающиеся порядком слагаемых или их знаками, как различные. Например,

$$2^2+1^2+1^2+1^2,$$
$$(-2)^2+(-1)^2+1^2+1^2,$$
$$1^2+2^2+(-1)^2+1^2$$

и т. д. считаются различными представлениями числа 7 в виде суммы четырех квадратов.

При таком соглашении можно показать, что если $r(m)$ — искомое число представлений числа m в виде суммы четырех квадратов, то

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

Проведя серию остроумных преобразований, Якоби доказал тождество

$$(1 + 2x^{1^2} + 2x^{2^2} + 2x^{3^2} + \ldots)^4 =$$
$$= 1 + 8\frac{x}{1-x} + 8\frac{2x^2}{1-x^2} + 8\frac{3x^3}{1-x^3} + 8\frac{5x^5}{1-x^5} + \ldots,$$

где степени l в выражениях $x^l/(1-x^l)$ пробегают все целые числа, *не делящиеся* на 4. Поскольку

$$\frac{x^l}{1-x^l} = x^l + x^{2l} + x^{3l} + \ldots,$$

легко убедиться в том, что $r(m)$ равно восьмикратной сумме всех не кратных четырем делителей числа m.

Поскольку единица является делителем любого числа и, очевидно, не кратна четырем, мы заключаем, что для всех $m > 0$

$$r(m) \geqslant 8,$$

откуда следует теорема Лагранжа (которая утверждает только, что $r(m) > 1$).

Тождество для рядов, которое привело к формуле Якоби для $r(m)$, — лишь одно из целого класса замечательных тождеств, открытых Якоби и другими. Некоторые из них связывают задачу partitio numerorum с проблемой представления чисел в виде суммы квадратов.

Открытие этих тождеств не было делом случая или особой удачи. Хотя в математике, как и всюду, открытия иногда обязаны счастливой случайности, везет обычно лишь достойным. Многие из этих тождеств были получены в результате систематического изучения класса аналитических функций, называемых *эллиптическими функциями*. В свою очередь идея эллиптических функций появилась у математиков при решении задач о нахождении периметра эллипса (отсюда эпитет «эллиптические») и об описании движения маятника. Хотя обе эти задачи приводят к интегралам в действительной области (относящимся к обычному интегральному исчислению), было замечено, что для их изучения гораздо больше подходит комплексная область. «Пересадка» проблемы из присущего ей окружения в, казалось бы, совершенно чужеродную ей область дала громадные преимущества, ибо только комплексный анализ сделал естественным и даже неизбежным удивительный переход от движения маятника к представлению целых чисел в виде суммы квадратов. Хотя многие из этих тождеств можно доказать прямыми и элементарными способами (например, не используя комплексные числа), однако обнаружить более глубокие причины того, «почему же они тикают», можно лишь обратившись к теории эллиптических функций. Тенденция к экономии мышления, обеспечиваемой использованием какой-либо более общей теории, столь же сильна среди математиков, как и среди других естествоиспытателей.

§ 10. Вероятность и независимость

Логические и исторические истоки теории (или исчисления) вероятностей связаны с простыми задачами подсчета. В эксперименте со случайным исходом (например, при вытягивании карты из колоды) вероятность осуществления данного события принимается равной отношению числа исходов, при которых это событие имеет место, к общему числу возможных исходов. Именно так подсчитывают шансы в азартных играх. В прошедшие времена теория вероятностей в основном и использовалась только для этого, однако в 20-м веке эта теория претерпела глубокие изменения и превратилась в важнейший и разветвленный раздел математики, нашедший применение не только в других ее разделах, но и в иных науках. Тем не менее логическая структура теории вероятностей удивительно проста. В этом параграфе мы дадим общее описание этой теории и обзор ее развития, а в следующем коснемся некоторых более специальных вопросов и результатов.

Исход многих экспериментов (таких, как бросание монеты или игральной кости, вытягивание карты из колоды и т. п.) нельзя с уверенностью предсказать заранее. Однако обычно можно перечислить *все возможные исходы*, и во многих случаях общее число таких исходов конечно (хотя иногда немыслимо велико). Допустим, что мы заинтересованы в осуществлении исхода, принадлежащего некоторому подмножеству исходов, и хотим каким-то разумным образом приписать событию, заключающемуся в том, что исход принадлежит этому подмножеству, некоторое число.

Более абстрактно, *каждому* подмножеству A множества Ω всех возможных исходов мы хотим приписать число $p(A)$, которое можно было бы считать мерой правдоподобности того, что исход принадлежит подмножеству A.

В случае, когда Ω — *конечное* множество, Лаплас предложил определять $p(A)$ как отношение числа $\nu(A)$ элементов A к общему числу $\nu(\Omega)$ элементов Ω, т. е.

$$p(A)=\frac{\nu(A)}{\nu(\Omega)},$$

разумеется, при условии, что все исходы, принадлежащие множеству Ω, можно считать равновероятными.

Последнее условие создает порочный круг, ибо понятие *вероятности* определяется через понятие *равновероятности*. С чисто технической точки зрения определение Лапласа сводит вычисление вероятностей к подсчетам.

Применение этого определения мы проиллюстрируем на типичном примере.

Допустим, что мы n раз подбрасываем монету и хотим найти вероятность того, что при этом герб выпадет ровно m раз ($1 \leqslant m \leqslant n$). Обозначим бросание, в результате которого выпал герб, буквой $Г$, а бросание, при котором выпала решетка, — буквой P. Тогда исход n бросаний можно записывать в виде последовательности n букв, каждая из которых есть либо $Г$, либо P. Назовем такую последовательность «словом». Тогда множество Ω всех возможных исходов n бросаний — это множество всех слов длины n, образованных из двух букв $Г$ и P. Можно показать, что общее число таких слов равно 2^n, т. е. что $\nu(\Omega) = 2^n$. Сколько из них содержат ровно m букв $Г$? Подсчет здесь сравнительно прост.

Проведем этот подсчет таким образом, чтобы проиллюстрировать один метод, очень полезный во многих ситуациях. Обозначим через $C(n; m)$ искомое число слов и рассмотрим $C(n+1; m)$, т. е. число слов длины $n+1$, содержащих m букв $Г$. Часть этих слов оканчивается буквой P; число их равно числу слов длины n, содержащих m букв $Г$. Остальные оканчиваются буквой $Г$; число их равно числу слов длины n, в которые буква $Г$ входит $m-1$ раз. Следовательно,

$$C(n+1;\ m) = C(n;\ m) + C(n;\ m-1), \qquad (*)$$

и мы получили то, что называется *рекуррентной формулой*; эта формула позволяет свести задачу для слов из $n+1$ букв к той же самой задаче для n-буквенных слов. Такой метод называется *индукцией*; с другим примером применения индукции (совсем иного характера) мы уже встречались при доказательстве леммы Шпернера.

Рекуррентная формула (*) позволяет найти для $C(n; m)$ явное выражение:

$$C(n;\ m) = \frac{n!}{(n-m)!\, m!}.$$

Числа $C(n; m)$ — это не что иное, как биномиальные коэффициенты:

$$(x+y)^n = C(n;\ 0)\, x^n + C(n;\ 1)\, x^{n-1}y + \ldots + \\ + C(n;\ m)\, x^{n-m}y^m + \ldots + C(n;\ n)\, y^n.$$

Если исходы всех n бросаний можно рассматривать как равновероятные, то по определению Лапласа вероятность m-кратного выпадения герба при n бросаниях монеты равна

$$p(m)=\frac{C(n;\ m)}{2^n}=\frac{1}{2^n}\,\frac{n!}{(n-m)!\,m!},$$

ибо общее число возможных исходов (т. е. n-буквенных слов из букв $Г$ и $Р$) равно 2^n, а число тех из них, в которых герб выпадает m раз, равно $C(n;\ m)$.

Допустим теперь, что мы n раз выбрасываем игральную кость и хотим узнать вероятность того, что двойка выпадет ровно m раз. Число всех возможных исходов здесь равно 6^n, и нам остается только подсчитать, в скольких из них двойка встречается ровно m раз.

Обозначим бросание, в котором выпала двойка, буквой $Г$, а остальные возможные исходы — символами P_1, P_3, P_4, P_5, P_6. Теперь на каждое слово длины n из двух букв $Г$ и $Р$, содержащее m букв $Г$ и $n-m$ букв $Р$, приходится 5^{n-m} слов той же длины (n) и с тем количеством (m) букв $Г$, где вместо $Р$ подставлены пять различных символов. Поэтому число предпочтительных исходов (т. е. таких слов, в которых ровно m раз встречается буква $Г$) равно

$$\frac{n!}{(n-m)!\,m!}\,5^{n-m}.$$

Поскольку, как замечено выше, общее число исходов равно 6^n, вероятность ровно m-кратного выпадения двойки при n бросаниях кости такова:

$$\frac{n!}{(n-m)!\,m!}\,\frac{5^{n-m}}{6^n}=\frac{n!}{(n-m)!\,m!}\left(\frac{1}{6}\right)^m\left(\frac{5}{6}\right)^{n-m}.$$

Допустим теперь, что наша монета неоднородна по весу (или изогнута), и вероятность выпадения герба при одном бросании равна $^1/_6$; тогда вероятность выпадения решетки при одном бросании равна $^5/_6$. Мы снова подбрасываем эту монету n раз и хотим узнать вероятность выпадения герба в точности m раз. Теперь уже довольно трудно описать равновероятные исходы, если не воспользоваться искусственным приемом, считая монету шестигранной костью, в которой одна грань соответствует гербу, а все другие — решетке. Ситуация становится еще неприятнее, если неоднородность монеты такова, что вероятность выпадения герба иррациональна, скажем равна $\sqrt{2}/2$. Приходится рассматривать многогранные кости и переходить к пределу, устремляя число граней к бесконечности.

Неудобство и логическая несостоятельность определения Лапласа способствовали подозрительному отношению математиков ко всему предмету теории вероятностей. Дело ухудшалось еще и тем, что попытки распространить определение Лапласа на случаи, когда число исходов бесконечно, приводили, казалось бы, к еще большим трудностям. Особенно плачевной стала ситуация после того, как Бертран попробовал найти вероятность того, что «случайно» выбранная хорда некоторого круга будет длиннее стороны правильного вписанного треугольника.

Если мы зафиксируем один конец хорды, то сможем считать окружность данного круга множеством Ω всех исходов, а дугу $\alpha\beta$ (рис. 6) — множеством A «предпочтительных исходов» (т. е. тех, в которых хорда длиннее стороны правильного вписанного треугольника). Поэтому представляется разумным полагать искомую вероятность равной $^1/_3$ (отношению длины дуги $\alpha\beta$ к длине окружности).

С другой стороны, можно считать, что хорда определяется ее серединой M, и рассматривать внутренность круга как множество Ω всех возможных исходов. Тогда множеством A «предпочтительных исходов» будет круг, заштрихованный на рис. 7; его радиус равен половине радиуса исходного круга. Теперь представляется не менее разумным считать искомую вероятность равной $^1/_4$, т. е. отношению площади заштрихованного круга к площади данного.

То, что два на вид вполне подходящих способа решения этой задачи привели к разным ответам, было настолько поразительным, что этот пример стали называть «парадоксом Бертрана». Конечно, это не логический парадокс, а попросту предостережение против бездумного употребления выражения «случайно». Вкупе с другими двусмысленными и неопределенными выражениями, этот термин немало способствовал усилению отрицательного отношения ко всему, имевшему что-либо общее со случаем и вероятностью.

Обсудив недостатки теории Лапласа, опишем и одно из ее выдающихся достижений.

Если неоднородность монеты такова, что вероятность выпадения герба при одном бросании равна p

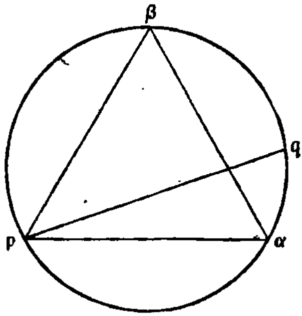


Рис. 6.

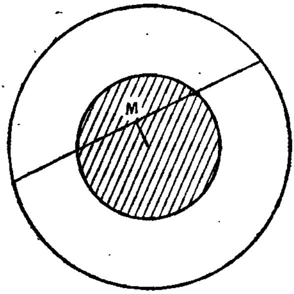


Рис. 7.

(и, следовательно, вероятность выпадения решетки равна $q = 1 - p$), то вероятность ровно m-кратного выпадения герба при n бросаниях составляет

$$\frac{n!}{m!\,(n-m)!}\,p^m q^{n-m}$$

(логические трудности мы пока оставляем без внимания). Лаплас (следуя более ранней работе Муавра) доказал следующую теорему: вероятность того, что число гербов, выпавших при n бросаниях, заключено между

$$np + \alpha\sqrt{2pqn} \quad \text{и} \quad np + \beta\sqrt{2pqn}$$

(где α и β — заданные числа), при возрастании n становится все ближе и ближе к величине интеграла

$$\frac{1}{\sqrt{2\pi}}\int_{\alpha}^{\beta} e^{-x^2/2}\,dx,$$

т. е. к площади области, ограниченной графиком функции $y=(1/\sqrt{2\pi})e^{-x^2/2}$, осью абсцисс и прямыми $x=\alpha$ и $x=\beta$ (рис. 8). Другими словами, при $n\to\infty$ вероятность того, что число гербов заключено между

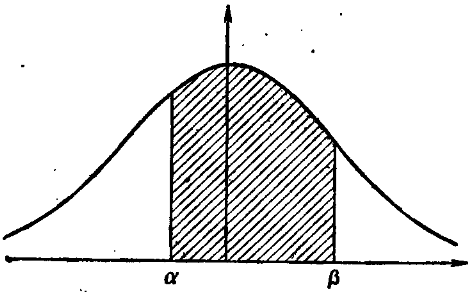


Рис. 8.

$np+\alpha\sqrt{2pqn}$ и $np+\beta\sqrt{2pqn}$, стремится к пределу

$$\frac{1}{\sqrt{2\pi}}\int_{\alpha}^{\beta} e^{-x^2/2}\,dx.$$

Чем же замечателен этот результат? Освобожденная от вероятностной терминологии теорема Лапласа (или точнее теорема Муавра — Лапласа) превратится в некоторое частное утверждение о биномиальных коэффициентах $C(n;m)$, но математическая литература и так кишит теоремами об этих коэффициентах. Чтобы оценить значение этой теоремы, нужно выйти за пределы чистой математики.

Во-первых, эта теорема не только *качественно* находится в согласии с нашей интуицией, но и уточняет интуицию *количественно*. «Вероятность есть уточненный здравый смысл» — так выразил это сам Лаплас.

Если мы будем подбрасывать «честную» (не фальшивую) монету ($p=q=1/2$) 10 000 раз, то в соответствии с интуицией мы будем ожидать приблизительно 5000 гербов. Теорема Муавра — Лапласа говорит нам, что с вероятностью около 0,99998 число гербов будет лежать между 4850 и 5150 и с вероятностью 0,8427 — между 4950 и 5050.

Во-вторых, кривая

$$y=\frac{1}{\sqrt{2\pi}}\,e^{-x^2/2}$$

или несколько более общая кривая

$$y=\frac{1}{\sigma\sqrt{2\pi}}\,e^{-(x-m)^2/2\sigma^2}$$

многократно встречалась в эмпирических исследованиях. Ее часто называют *нормальной* кривой, или кривой нормального распределения. То, что нашлась *математическая модель*, которая привела к этой кривой, было весьма знаменательно.

История теории вероятностей дает нам прекрасный пример взаимоотношения двух противоположных точек зрения на математику — исходящей из потребностей «чистой» и из потребностей «прикладной» науки. Для пуриста теорема Муавра — Лапласа является, самое большее, вкладом в широкое поле узкого раздела знания, касающегося биномиальных коэффициентов. Первоначальное доказательство основывалось на известной асимптотической формуле Стирлинга

$$n!\sim\left(\frac{n}{e}\right)^n\sqrt{2\pi n},$$

где символ $\sim$ означает, что имеет место предельное соотношение

$$\lim_{n\to\infty}\left[n!\Big/\left(\frac{n}{e}\right)^n\sqrt{2\pi n}\right]=1.$$

Поэтому пурист рассудил бы, что если этой теореме и можно придать какую-то «глубину», то исключительно по милости Стирлинга. Наконец, он отверг бы вероятностную интерпретацию по причине ее логической несостоятельности и остался бы глух к тому соображению, что теорема Муавра — Лапласа, несомненно, оказалась важной ступенью на пути исследования широкого многообразия случайных явлений.

Столь суровое отношение к вероятностным законам действительно существовало, и оно привело к тому, что теория вероятностей как математическая дисциплина временно зачахла и интерес к ней возродился только в начале 20-го века благодаря успешным и эффектным ее применениям в физике.

В ретроспективе логические затруднения, с которыми столкнулась теория Лапласа, оказались незначительными; тем не менее попытки дать строгое обоснование теории вероятностей исключительно благотворно повлияли на предмет в целом.

Современная точка зрения совсем проста. Из множества всех возможных исходов (называемого «пространством выборок») выбирается совокупность подмножеств (называемых «элементарными событиями»), которым раз и навсегда приписываются вероятности. Вероятности более сложных событий вычисляются при помощи двух следующих аксиом.

1. Аксиома аддитивности. Если E_1, E_2, ... — несовместимые события (т. е. соответствующие им подмножества в пространстве выборок не имеют общих элементов), то вероятность (Prob) события $\{E_1$ или E_2 или $\ldots\}$ равна сумме вероятностей составляющих событий (разумеется, при условии, что этим составляющим событиям можно приписать вероятности); итак,

$$\text{Prob}\{E_1 \text{ или } E_2 \text{ или } \ldots\} = \text{Prob}\{E_1\} + \text{Prob}\{E_2\} + \ldots .$$

2. Аксиома дополнения. Если некоторому событию E можно приписать определенную вероятность, то и событию «не E» можно приписать вероятность.

Наконец, всему пространству выборок приписывается вероятность 1:

$$\text{Prob}\{\Omega\} = 1.$$

Воспользовавшись обеими аксиомами, можно, например, подсчитать, что если Prob $\{E\}$ определена, то

$$\text{Prob}\{\text{не } E\} = 1 - \text{Prob}\{E\}.$$

Почему выбраны именно эти аксиомы? Обычно от аксиом требуется, чтобы они сводили в некий кодекс интуитивные представления и допускали непосредственную *проверку* в ряде простых случаев.

Приведенные выше аксиомы, очевидно, выполняются во всех ситуациях, в которых определение Лапласа применимо и не приводит к неопределенности, и находятся почти в полном соответствии с интуитивным представлением о вероятностях.

Важное исключение из этого правила мы находим в квантовой механике. Допустим, что за экраном (a) с двумя малыми отверстиями A и B помещен источник S моноэнергетических электронов (рис. 9). Электроны, пройдя через отверстия, достигают

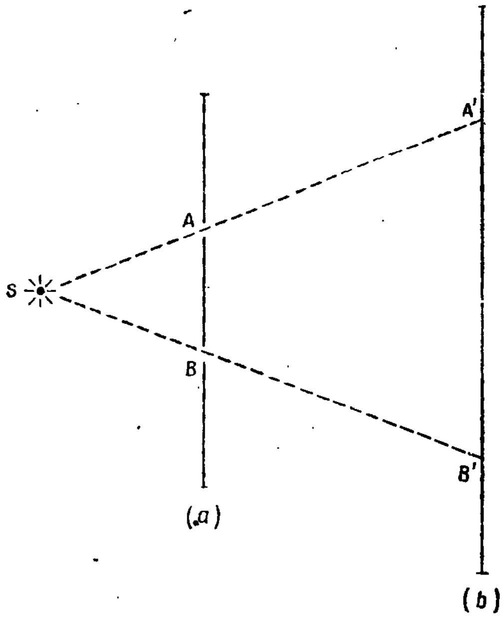


Рис. 9.

другого экрана (b), где они регистрируются счетчиками. Допустим, что отверстие B закрыто. Тогда электроны могут проходить только через A, и если бы они вели себя строго классическим образом, они собирались бы вблизи точки A', в которой прямая, соединяющая S с центром отверстия A, пересекает экран (b). На самом деле их расположением управляет «случай», и лучшее, что можно сделать, — это приписать каждой области R на экране (b) вероятность $P_A(R)$ того, что электрон попал внутрь R. Аналогично, $P_B(R)$ — вероятность того, что электрон, вылетевший из S, попадет в область R на экране (b) при условии, что закрыто отверстие A.

Если оба отверстия открыты, то в соответствии с аксиомой аддитивности вероятность $P_{A \text{ или } B}(R)$ должна была бы равняться сумме $P_A(R) + P_B(R)$. Однако экспериментально установлено, что это не так:

$$P_{A \text{ или } B}(R) \neq P_A(R) + P_B(R).$$

Таким образом, кажущееся очевидным утверждение, что электрон, попавший на экран (b), *должен* был пройти либо через A, либо через B, несостоятельно. Попытка установить через какое отверстие на самом деле прошел электрон связана со столь сильным вмешательством в ход эксперимента, что его конечный результат изменяется и аксиома аддитивности восстанавливается!

Аксиомы аддитивности и дополнения являются слишком общими и всеобъемлющими, чтобы они одни могли служить основанием такой богатой и плодотворной теории, как теория вероятностей. Тем не менее, как мы увидим в следующем параграфе, две эти аксиомы имеют глубокое математическое содержание.

В основе предмета лежит выбор «элементарных» событий и решение вопроса о том, какие вероятности им следует приписать. Здесь вступают в силу нематематические соображения, и нам остается лишь уповать на то, что эмпирический мир откроет нам путь в многообещающие области исследования.

Вернемся к бросаниям монеты. Пытаясь построить какую-нибудь реалистическую и полезную теорию, мы должны сначала рассмотреть два совершенно различных вопроса:

(1) Какова монета, которую бросают?

(2) Каков механизм бросаний?

Первый вопрос касается свойств самой монеты; второй — того, имеется или нет какая-нибудь зависимость между последовательными бросаниями (и если да, то какая).

Чтобы полнее оценить значение этих двух моментов, приведем такой пример. Допустим, что рассматривается статистическая структура английского языка. Буквы в английских текстах встречаются с определенной частотой, которая мало меняется от текста к тексту. Так, эмпирически установлено, что около 13,05% всех букв падает на букву e, 9,02% — на t, 6,81% — на a. Вообразим 26-гранную кость, на гранях которой написаны буквы алфавита, причем допустим, что эта кость неоднородна по весу, так что вероятность выпадения каждой грани равна частоте, с которой написанная на ней буква встречается в языке. Можно даже добавить 27-ю грань, отвечающую знаку пробела между словами, и сделать соответствующее перераспределение весов. Бросая эту кость, скажем, 10 000 раз, мы получим текст, в котором буквы и пробел будут встречаться почти с той же частотой, что и в английском языке. Однако это не будет похоже на английский текст, пока мы не позаботимся о том, чтобы определенным образом *скоррелировать* бросания. В самом деле, мы знаем, что после t чаще всех встречается буква h, что n чаще всех других букв бывает в конце слова (т. е. предшествует пробелу) и т. д. С учетом этих уточнений полученный текст станет намного больше похожим на английский, а если еще навести порядок с триграммами (т. е. наборами из трех последовательных букв), то текст уже вполне можно будет выдавать за английский.

Эксперименты такого рода придумал и осуществил Клод Шеннон в связи со своими первыми в этой области прекрасными работами по теории информации.

Вернемся к нашей монете. Простейшие предположения здесь состоят в том, что монета «честная»; выпадению как герба, так и решетки приписывается вероятность $^1/_2$, а бросания считаются независимыми.

Понятие *независимости* играет центральную роль в теории вероятностей, поэтому мы обсудим его несколько подробнее.

События E и F независимы в обычном смысле слова, если осуществление одного из них никак не влияет на осуществление другого.

В таком случае мы должны иметь возможность вычислять вероятность составного события $\{E$ и $F\}$, если известны вероятности составляющих E и F.

Иными словами, если E и F независимы, то должно существовать *правило*, позволяющее вычислить Prob $\{E$ и $F\}$ при условии, что известны Prob $\{E\}$ и Prob $\{F\}$. Более того, это правило должно быть *универсальным*, т. е. применимым к любой паре независимых событий.

Такое правило имеет вид функции $f(x, y)$ от двух переменных x и y, и, подытоживая предыдущее, мы можем сказать, что если E и F независимы, то

$$\text{Prob}\,\{E \text{ и } F\} = f\,(\text{Prob}\,\{E\},\ \text{Prob}\,\{F\}).$$

Рассмотрим теперь следующий эксперимент. Представим себе монету, которую можно «испортить» любым требуемым образом (т. е. сделать вероятность p выпадения герба равной любому числу между 0 и 1), а также четырехгранную кость, обладающую тем же свойством. Пусть грани кости обозначены цифрами 1, 2, 3, 4, а соответствующие вероятности равны p_1, p_2, p_3, p_4, причем все p_i неотрицательны и $p_1 + p_2 + p_3 + p_4 = 1$. Что бы ни означала независимость, естественно ввести предположение, что эту монету и эту кость можно бросать независимо. Рассмотрим, например, событие $\{Г$ и (1 или 2)$\}$. Тогда, с одной стороны,

$$\text{Prob}\,\{Г \text{ и } (1 \text{ или } 2)\} = f\,(p,\ p_1 + p_2);$$

с другой стороны, поскольку событие $\{Г$ и (1 или 2)$\}$ эквивалентно событию $\{(Г$ и 1) или ($Г$ и 2)$\}$, мы имеем

$$\text{Prob}\,\{Г \text{ и } (1 \text{ или } 2)\} = \text{Prob}\,\{Г \text{ и } 1\} + \text{Prob}\,\{Г \text{ и } 2\} =$$
$$= f\,(p,\ p_1) + f\,(p,\ p_2).$$

(Заметим, что мы повторно использовали аксиому аддитивности.) Таким образом.

$$f\,(p,\ p_1 + p_2) = f\,(p,\ p_1) + f\,(p,\ p_2)$$

для всех p, p_1, p_2, выбор которых ограничен лишь неравенствами

$$0 \leqslant p \leqslant 1,\quad 0 \leqslant p_1,\quad 0 \leqslant p_2,\quad p_1 + p_2 \leqslant 1.$$

Если считать (что представляется вполне разумным) функцию f непрерывно зависящей от своих аргументов, то из предыдущего равенства вытекает, что $f(x, y) = xy$; следовательно, вероятность совместного осуществления независимых событий должна равняться *произведению* вероятностей каждого из них в отдельности.

Это рассуждение (которым мы обязаны Г. Штейнгаузу) является прекрасной иллюстрацией неформальных (можно было бы сказать «закулисных») соображений, предшествующих формальному определению. Это рассуждение такого рода: «Мы на самом деле не знаем, что такое независимость, но чем бы она ни была, если она имеет смысл, то она должна обладать следующими свойствами...». Получив из этих свойств подходящие следствия (например, $f(x, y) = xy$), математик уже может установить логическую связь понятий и предложить *формальное определение*.

Два события E и F (или любое конечное число событий) называются *независимыми*, если к ним применимо *правило умножения вероятностей*, т. е. если

$$\mathrm{Prob}\{E \text{ и } F\} = \mathrm{Prob}\{E\} \cdot \mathrm{Prob}\{F\}.$$

Существует еще один способ обосновать правило умножения вероятностей для независимых событий. Он опирается на понятие частоты. Допустим, что в n испытаниях $n(E)$ раз происходит событие E, $n(F)$ раз — событие F и $n(E \text{ и } F)$ раз происходят одновременно оба события E и F. Интуитивно кажется, что частота, с которой событие встречается в длинной серии испытаний, должна в некотором смысле служить приближением его вероятности[1]. Поскольку

$$\frac{n(E \text{ и } F)}{n} = \frac{n(E \text{ и } F)}{n(F)} \cdot \frac{n(F)}{n},$$

мы видим, что частота одновременного осуществления событий E и F в n испытаниях равна произведению частоты события E в $n(F)$ испытаниях, в которых осуществляется F, на частоту события F.

Теперь можно было бы рассуждать так: если E и F независимы, то информация о том, что произошло событие F, не позволяет предсказать осуществление или неосуществление события E; поэтому частота события E в испытаниях, в которых осуществляется F, должна приближенно равняться общей частоте события E в n испытаниях. Иначе говоря, можно ожидать, что с некоторой степенью точности

$$\frac{n(E \text{ и } F)}{n(F)} = \frac{n(E)}{n}$$

---

[1]) Физики (проявляя некоторую логическую нетребовательность) фактически отождествляют вероятности с частотами. Такой подход связан со многими трудностями, ибо когда число испытаний неограниченно возрастает, приходится переходить к пределам. Рихард фон Мизес пытался аксиоматизировать теорию вероятностей, опираясь на рассмотрение частот в определенных *бесконечных* последовательностях испытаний, которые он называл *коллективами*. С некоторыми возникшими здесь на первых порах логическими трудностями удалось справиться А. Вальду, однако этот подход не получил сколько-нибудь широкой поддержки среди математиков.

и, следовательно,

$$\frac{n(E \text{ и } F)}{n} = \frac{n(E)}{n} \cdot \frac{n(F)}{n};$$

это сильно напоминает предыдущее соотношение

$$\text{Prob}\{E \text{ и } F\} = \text{Prob}\{E\} \cdot \text{Prob}\{F\}.$$

Такое эвристическое «обоснование» ничего не доказывает, но оно возвращает нас к правилу умножения вероятностей, которое выступает здесь в совершенно другом контексте[1]).

Договорившись о том, что независимость означает применимость правила умножения вероятностей, вернемся к лапласовой монете.

Если бросания предполагаются независимыми, то вероятность, соответствующая некоторому конкретному слову вида

$$ГГРР \ldots РГ,$$

содержащему m букв $Г$ (и $n-m$ букв $Р$), равна $p^m q^{n-m}$. Всего имеется $C(n; m)$ таких слов, поэтому (в соответствии с аксиомой аддитивности) вероятность того, что при n *независимых* бросаниях герб выпадет ровно m раз, равна $C(n; m) p^m q^{n-m}$. Мы пришли к той же формуле, что и раньше, но уже без ссылок на «равновероятность»; вместо этого мы опирались на понятие независимости.

И это не просто перевод рассуждения на другой язык: анализ нормальной кривой

$$y = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$

[1]) Хотя формально ничего не доказано, наше ощущение, что мы на правильном пути, несколько укрепилось, ибо нам удалось как-то «свести концы с концами». В теоретической физике возможность такого «увязывания» является важным (а часто и единственным) путем к истине. В математике же какая-то незамеченная противоречивость или ошибочность предпосылки могут создать иллюзию, что все в порядке, в то время как на самом деле порок рассуждения неустраним, и остается только вместе с Эйнштейном надеяться на то, что «raffiniert ist der Herr Gott, aber boshaft ist er nicht» («господь бог изощрен, но не злонамерен»).

показал, что она является результатом прежде всего независимости испытаний, а не магической силы формулы Стирлинга.

Рассмотрим, например, последовательность по-разному испорченных монет, которые подбрасывают независимо одну от другой (иногда этот эксперимент называют схемой Пуассона). Пусть p_k и $q_k = 1 - p_k$ — вероятности выпадения соответственно герба и решетки для k-й монеты. В этом случае не существует простой и компактной формулы, выражающей вероятность того, что при n независимых бросаниях герб выпадет ровно m раз. Однако, приспособив к этой задаче метод производящих функций, который обсуждался в § 7, можно доказать следующие обобщения теоремы Муавра — Лапласа.

Вероятность того, что число выпадений герба при n бросаниях лежит между

$$(p_1 + \ldots + p_n) + \alpha \sqrt{p_1 q_1 + \ldots + p_n q_n}$$

и

$$(p_1 + \ldots + p_n) + \beta \sqrt{p_1 q_1 + \ldots + p_n q_n}$$

(где α и β — заданные фиксированные числа), при $n \to \infty$ стремится к интегралу

$$\frac{1}{\sqrt{2\pi}} \int_{\alpha}^{\beta} e^{-x^2/2}\, dx,$$

если бесконечный ряд $p_1 q_1 + p_2 q_2 + \ldots$ расходится[1]).

Если отвлечься от довольно слабого условия, касающегося сходимости ряда, нормальный закон, выражаемый кривой

$$y = \frac{1}{\sqrt{2\pi}} e^{-x^2/2},$$

представляется весьма универсальным, по крайней мере если речь идет о независимых испытаниях или

[1]) Заметим, что это условие совершенно естественно. Если бы этот ряд сходился, то его общий член $p_n q_n$ становился бы очень малым при больших n, т. е. монеты с большими номерами были бы настолько испорчены «в пользу» герба или решетки, что на «законы случая» налагались бы слишком сильные ограничения.

событиях. В 20-х и 30-х годах нашего века было выяснено, в какой мере он универсален. Соответствующие открытия слишком сложны технически, и говорить о них здесь невозможно; однако в процессе этих исследований были обнаружены многие факты, оказавшиеся ценными для других разделов математики и иных наук. Это — та награда, которую заслужил Лаплас за свою веру в важность теории вероятностей.

В настоящее время теорема Муавра — Лапласа и ее обобщение на схемы Пуассона являются лишь частными следствиями весьма общей *центральной предельной теоремы*. Однако, как и все другие великие теоремы, эти частные случаи содержали почти все зерна той общей истины, которая в конце концов их поглотила.

Понимание того, что в основе нормального закона лежит независимость, позволяет применять теоремы, подобные теореме Муавра — Лапласа, в ситуациях, не имеющих ничего общего со случайными явлениями.

Рассмотрим, например, натуральные числа 1, 2, 3, 4, ... и простые числа 2, 3, 5, 7, 11, ..., которые мы будем обозначать $P_1, P_2, \ldots$ (поскольку символы $p_1, p_2, \ldots$ используются здесь для обозначения вероятностей); так, $P_1 = 2$, $P_2 = 3$, ...

Пусть E — некоторое множество целых чисел. Обозначим через $K_n(E)$ число элементов множества E среди первых n натуральных чисел 1, 2, ..., n. Если при $n \to \infty$ существует[1]) предел

$$D\{E\} = \lim_{n\to\infty} \frac{K_n(E)}{n},$$

то мы назовем его *плотностью* множества E.

Пусть E_i — множество целых чисел, делящихся на i-е простое число P_i. Довольно ясно, что

$$D\{E_i\} = 1/P_i.$$

---

[1]) Следует заметить, что установление существования некоторого предела часто бывает неизмеримо более трудной частью теоремы, чем его фактическое вычисление. Например, в теореме о простых числах, упомянутой в § 1, глубокий результат состоит в том, что предел

$$\lim_{n\to\infty} \frac{\pi(n)}{n/\log n}$$

существует; как только это установлено, вывести, что этот предел равен 1, уже не составляет никакого труда.

Рассмотрим далее множество $E_1 \cap E_2 \cap \ldots \cap E_r$ целых чисел, делящихся на первые r-простых чисел $P_1, P_2, \ldots, P_r$. (Символ $\cap$ здесь означает «пересечение» множеств, т. е. $E_1 \cap E_2 \cap \ldots \cap E_r$ есть множество элементов, общих всем множествам $E_1, E_2, \ldots, E_r$.) Тогда

$$D\{E_1 \cap E_2 \cap \ldots \cap E_r\} = 1/(P_1P_2 \ldots P_r) = D\{E_1\} D\{E_2\} \ldots D\{E_r\},$$

и нетрудно заметить аналогию с правилом умножения вероятностей.

Эта аналогия приводит нас к попытке применить в этой ситуации (быть может, с некоторой натяжкой) обобщение Пуассона теоремы Муавра — Лапласа, т. е. заподозрить, что плотность множества целых чисел n, для которых число простых делителей $\nu(n)$ заключено между $\log\log n + \alpha\sqrt{\log\log n}$ и $\log\log n + \beta\sqrt{\log\log n}$, равна интегралу

$$\frac{1}{\sqrt{2\pi}} \int_{\alpha}^{\beta} e^{-x^2/2}\,dx.$$

[Вот примеры числа $\nu(n)$: $\nu(20) = \nu(2^2 \cdot 5) = 2$, $\nu(90) = \nu(2 \cdot 3^2 \cdot 5) = 3$, $\nu(13) = 1$.] Итак, гипотеза состоит в том, что

$$D\{\log\log n + \alpha\sqrt{\log\log n} < \nu(n) < \log\log n + \beta\sqrt{\log\log n}\} = \frac{1}{\sqrt{2\pi}} \int_{\alpha}^{\beta} e^{-x^2/2}\,dx.$$

Эта гипотеза подтвердилась: доказано, что число простых делителей действительно распределено в соответствии с нормальным законом.

Здесь перед нами пример того, как понятия и методы, созданные в одной области, находят важное приложение в другой, весьма от нее далекой. Нормальный закон, который обычно мысленно связывают с вероятностью и случайностью, оказывается применимым в теории чисел, в той ветви чистой математики, которая по праву считается наиболее строгой и наименее «случайной»...

§ 11. Мера

Проблема измерения величин областей на плоскости и в трехмерном пространстве, состоящая в том, что этим областям нужно приписать числа, выражающие их площади или объемы, восходит к самым истокам математики. Греки создали стройную теорию площадей многоугольников и объемов многогранников. Интегральное исчисление (которое берет начало в некоторых исследованиях Архимеда) распространило эту теорию на широкий класс областей, ограниченных кривыми линиями и искривленными поверхностями.

Однако в процессе дальнейшего развития математики этого оказалось недостаточно: понадобилось приписать меры множествам более широкого класса. Так появилась общая теория меры, разработанная Борелем и Лебегом. Но и это обобщение не достаточно полно: существуют множества, которым нельзя приписать меру Лебега (т. е. эти множества неизмеримы). Чтобы построить неизмеримые множества, приходится применять знаменитую аксиому выбора, согласно которой из любой заданной совокупности непересекающихся множеств можно образовать новое множество, выбрав по одному элементу из каждого множества исходной совокупности. Хотя эта аксиома звучит вполне безобидно, многие ее следствия могут показаться странными и парадоксальными.

Теория меры имеет важные применения в теории вероятностей; об этом пойдет речь в § 12.

Многие математические идеи, которые сейчас кажутся принадлежащими алгебре или анализу, обязаны своим происхождением задачам, возникшим из геометрии. Так обстоит дело и с понятием меры и измерения.

К рассмотрению *длины* отрезка прямой или дуги кривой, площади или объема области привели, вероятно, самые первые попытки расширить рамки использования чисел, которое ранее сводилось лишь к пересчету отдельных предметов. Эти понятия в своем первоначальном виде появились уже в самых ранних математических работах вавилонян, египтян, греков.

Евклид занимался главным образом площадями многоугольников и объемами многогранников. При этом использовались две аксиомы:

1. Если многоугольники (многогранники) конгруэнтны, то их площади (объемы) равны.

2. Если многоугольник (многогранник) допускает разбиение на конечное число неперекрывающихся многоугольников (многогранников) $A_1, A_2, \ldots, A_n$, то

его площадь (объем) равна сумме площадей (объемов) составляющих $A_1, A_2, \ldots, A_n$ [1]).

Если, кроме того, выбран некоторый определенный квадрат (соответствеино куб), площадь (соответственно объем) которого приняты за единицу, то каждому многоугольнику (многограннику) можно однозначно приписать число, выражающее в выбранных единицах его площадь (объем).

Однако даже для нахождения площади круга приходится выйти из безопасной области коиечного и ввести в полной мере бесконечную операцню перехода к пределу. В древности самый большой вклад в нахождение площадей и объемов фигур, ограниченных кривыми линиями и искривленными поверхностями, внес Архимед; есть все основания полагать, что Архимед полностью осознавал всю тонкость понятия предела. Эти исследования Архимеда, без сомнения, явились началом интегрального исчисления, хотя вычнсление площадей и объемов было в полной мере систематизировано лишь во времена Ньютона и его последователей.

Интегральное исчисление позволяет приписать площади, объемы и длины лишь сравнительно «смирным» множествам. Оно не дает никаких средств измерять, например, такие множества, как совокупность всех точек (x, y) с рациональными координатами, заключенными между 0 и 1. Но вопрос об измерении таких множеств н не поднимался, пока проблематика интегрального исчисления отражала, в основном, нужды физики и геометрии.

В конце 19-го века возникли задачи, которые привели к необходимости приписать численную меру значительно более широкому классу множеств. Возросший интерес к вопросам сходимости и расходимости привлек внимание к множествам сходимости и расходимости и к проблеме определения их «разме-

[1]) Пурист признает такое разбиение невозможным, поскольку два смежных многоугольнка (многогранника) должны иметь общую часть граинцы; если же считать по определению, что многоугольники (многогранники) не содержат своих границ (т. е. «открыты»), то они не смогут исчерпать A.

ров». Канторовская теория множеств, ставшая краеугольным камнем всей современной математики, возникла в результате изучения Кантором тригонометрических рядов и их множеств сходимости.

Проблема меры формулируется совсем просто. Мы хотим уметь приписывать некоторому множеству A *неотрицательное число* $m(A)$, называемое его *мерой*, так, чтобы выполнялись следующие условия:

1. Если $A_1, A_2, \ldots$ — непересекающиеся *измеримые* множества (т. е. каждому A_i можно сопоставить его меру $m(A_i)$), то и их *объединение* $A_1 \cup A_2 \cup \ldots$ (т. е. множество, состоящее из элементов всех множеств $A_1, A_2, \ldots$) измеримо; при этом

$$m(A_1 \cup A_2 \cup \ldots) = m(A_1) + m(A_2) + \ldots .$$

2. Если A и B — измеримые множества и A содержится в B ($A \subset B$), то их разность $B - A$ (т. е. множество всех точек из B, *не принадлежащих* A) тоже измерима. Тогда, согласно свойству 1, $m(B - A) = m(B) - m(A)$.

3. Некоторое множество E считается имеющим меру 1 (единичное множество):

$$m(E) = 1.$$

4. Если два множества конгруэнтны, то их меры равны (при условии, что множества измеримы).

При изучении множеств точек на прямой за E принимают некоторый отрезок, на плоскости — квадрат, в пространстве — куб. Такой выбор диктуется желанием, чтобы меры, приписываемые «смирным» множествам, совпадали с мерами, сопоставленными им раньше в геометрии или в интегральном исчислении.

Можно ли существенно расширить класс измеримых множеств, если приписывать меры в соответствии с перечисленными правилами? Ответом служит уверенное «да», при условии (в этой ситуации решающем), что в правиле 1 допускается *бесконечное число* множеств A_i.

Если, следуя Евклиду, допустить лишь конечное число множеств A_i (в этом случае мера называется *конечно аддитивной*), мы почти ничего не выиграем, и расширение класса «смирных» множеств будет совсем незначительным. Например, можно показать, что множество точек единичного квадрата с рациональными координатами окажется неизмеримым, т. е. ему нельзя будет приписать никакой меры, не впадая при этом в противоречие.

Ситуация коренным образом изменится, если рассматривать *вполне аддитивную* меру, т. е. допускать в правиле 1 бесконечное число множеств A_i. В этом случае класс измеримых множеств существенно вырастет и почти все множества, использовавшиеся в классической и возникшие в современной математике, окажутся измеримыми.

Вполне аддитивная мера была введена в начале 20-го века Эмилем Борелем и Анри Лебегом [1]) и стала источником наиболее строгой и плодотворной линии исследования в математике. Мера Лебега — одно из самых мощных средств современного анализа.

Насколько же общей является эта мера? Всякое ли множество на прямой измеримо? Витали первым показал, что даже мере Лебега свойственны определенные ограничения, т. е. существуют множества точек, для которых она не определена. Поясним это, рассматривая для простоты множество на окружности, а не на прямой; существо дела от этого не меняется. (Два множества на окружности конгруэнтны, если они совмещаются при некотором повороте этой окружности.) Определим на окружности множество точек (углов) Z, которому нельзя приписать меру так, чтобы выполнялись перечисленные выше условия. Пусть x — некоторая точка окружности. Рассмотрим все точки, которые получаются из x поворотом окружности на углы, составляющие рациональную часть от 2π. Эти углы образуют счетное множество; обозначим их $\alpha_1, \alpha_2, \ldots, \alpha_n, \ldots$. Множество *всех* углов круга можно разбить на попарно непересекающиеся классы C, в каждый из которых попадают «рацио-

[1]) Кроме перечисленных выше четырех правил, Лебег постулировал еще одно: всякое подмножество множества меры нуль измеримо (и, следовательно, имеет меру нуль).

нально сравнимые» углы, т. е. углы, отличающиеся друг от друга на рациональные величины. Эти классы не имеют общих элементов. Выберем теперь из каждого класса в точности одну точку и обозначим полученное множество через Z. Поворачивая затем множество Z на углы α_1, α_2, ..., мы получим новые множества, причем *все они попарно не пересекаются.* Совокупность всех точек всех этих множеств есть множество всех точек окружности. Кроме того, количество этих множеств счетно и все они конгруэнтны, ибо получаются одно из другого поворотом на углы α_1, α_2, ... Итак, мы получили счетную совокупность множеств Z, Z_1, ..., Z_n, ..., которые попарно не пересекаются, конгруэнтны и их объединением является вся окружность, которую мы можем выбрать в качестве единичного множества E, $m(E) = 1$. Какова же мера множества Z? Допустим, что она равна 0. Тогда и любое другое множество этой совокупности имеет меру 0, а их счетное объединение должно иметь меру 1, так как $m(E) = 1$. Следовательно, в этом случае нарушается аксиома аддитивности. Если же мера множества Z положительна, то мы получим сумму бесконечного числа одинаковых положительных слагаемых; такая сумма не может равняться 1, что снова является нарушением аксиомы аддитивности.

Итак, мы указали множество, которому нельзя приписать меру. Это еще одно доказательство *невозможности*, свидетельствующее о том, что процесс обобщения, как везде в математике, должен продолжаться. Однако при определении множества Витали мы воспользовались крайне неконструктивным приемом. В самом деле, мы *выбирали* по одному элементу из каждого класса C. Но как? Классы C определены слишком неявно, чтобы можно было указать явное правило выбора из этих элементов. И все же нас не покидает ощущение, что такой выбор элемента из каждого класса и объединение этих элементов в новое множество должны быть дозволены, даже если мы неспособны дать конкретное предписание, как выполнить эту задачу.

Выход из этой дилеммы указал в начале 20-го века Эрнест Цермело. Он предложил узаконить «построения», подобиые описанному выше, прняяв следующую общую аксиому:

Если задана совокупность C непересекающихся множеств, то можно образовать множество Z, *выбрав* по одному элементу из каждого множества этой совокупности.

Эта аксиома известна под названием *аксиомы выбора.* С самого своего рождения она была предметом горячих дебатов, поскольку многие ее следствия выглядят весьма странными и парадоксальными.

Так, например, Банах и Тарский доказали, что две сферы S_1 и S_2 *различных* радиусов можно разбить на одно и то же конечное число попарно не пересекающихся множеств:

$$S_1 = A_1 + A_2 + \ldots + A_n,$$
$$S_2 = B_1 + B_2 + \ldots + B_n,$$

так, чтобы A_i было конгруэнтно B_i при любом $i = 1, 2, \ldots, n$. Этим множествам нельзя приписать никакие меры, ибо, составляя их вместе однным способом, мы получаем бо́льшую сферу, а используя другое пространственное расположение — меньшую! Между прочим, на плоскости такое разбиение невозможно. Как показал Банах, для всех подмножеств плоскости можио найтн *конечно аддитивную меру,* обладающую тем свойством, что меры конгруэнтных множеств равны.

Попытки обобщить меру Лебега были вызваны необходимостью. Как мы уже говорили, математики включали в рассмотрение множества все более и более общего вида. Например, в теории тригонометрических рядов можно было сформулировать теоремы, верные для всех действительных чнсел, *за исключением* некоторого специального множества. При этом желательно было установить, пользуясь какими-то строго определенными понятиями, что множество этих исключительных точек в некотором смысле мало и им можно пренебречь. В несчетном континууме точек можно «пренебрегать» счетными множествами, считая их малыми, однако в большинстве случаев исключительные множества оказываются несчетными, но имеющими лебегову *меру нуль*. Многие утверждения теории вероятностей выполняются «с вероятностью единица» (или «почти достоверно»). Это попросту означает, что они верны для «почти всех» точек некоторого подходящего множества, т. е. для всех его точек, за исключением множества меры нуль. Некоторые важные теоремы статистической механики устанавливают свойства динамических систем, верные лишь для *почти всех* начальных условий.

Одно последнее замечание. Понятие меры находится в полном согласии с самой примитивной интуицией. Аксиома выбора, всего-навсего позволяющая рассматривать множество Z элементов, выбранных по одному из каждого множества некоторого семейства непересекающихся множеств, звучит настолько естественно, что кажется почти тривиальной. И тем не менее эта аксиома приводит к парадоксу Банаха — Тарского!

Вот почему оказался совершенно необходимым критический пересмотр логических оснований теории множеств, а вопрос существования математических «созданий» стал серьезной проблемой.

Если, как заявил Пуанкаре, существовать — это только быть свободным от противоречия, то у нас нет другого выхода, кроме как научиться уживаться с неприятностями вроде неизмеримых множеств или разбиений Банаха — Тарского.

§ 12. Еще о теории вероятностей

Читатель, наверное, заметил, что аксиомы аддитивности и дополнения, лежащие в основании теории вероятностей, *совпадают* с аксиомами меры. Следовательно, вероятности — это не что иное, как меры, а теория вероятностей — раздел теории меры.

Осознание этого обстоятельства, самого по себе довольно тривиального, имело серьезные и весьма глубокие последствия. Во-первых, благодаря этому теория вероятностей была введена в рамки математической строгости и стала «респектабельной». Во-вторых, что гораздо важнее, это сильно раздвинуло ее границы и позволило ставить и решать совершенно новые задачи.

В § 9 мы не уточняли, конечное или бесконечное множество событий имеется в виду в аксиоме аддитивности. В настоящее время принято допускать бесконечное множество событий, т. е. рассматривать *счетно аддитивную* меру, ибо это позволяет обратиться к задачам, связанным с бесконечными последовательностями испытаний.

Счетная аддитивность нужна даже в самых простых случаях. Если, например, два человека A и B по очереди бросают каждый свою монету (сначала бросает A, потом B, затем снова A и т. д.), то можно ставить вопрос о вероятности того, что первым, у кого выпадет герб, окажется A. Это может случиться либо при первом бросании, либо при третьем (если первые два раза выпадет решетка), либо при пятом и т. д. Таким образом, рассматриваемое событие разлагается на бесконечное число несовместимых событий. Если монеты «честные», а бросания независимы, то вероятности составляющих событий равны

$$\frac{1}{2}, \frac{1}{2^3}, \frac{1}{2^5}, \ldots,$$

и мы получаем искомую вероятность в виде

$$\frac{1}{2}+\frac{1}{2^3}+\frac{1}{2^5}+\ldots=\frac{2}{3},$$

при условии, что применима аксиома счетной аддитивности.

Рассмотрим теперь такую задачу. Допустим, что мы долго бросаем «честную» монету, причем отдельные бросания независимы. Какова частота выпадения герба? Интуиция подсказывает нам, что в ответе должна получиться 1/2. Но что это означает? Конечно, даже при очень большом числе бросаний глупо было бы ожидать, что *точно* в половине случаев выпадет герб, а в другой половине решетка. Ясно, что имеет смысл искать формулу, которая выполняется лишь *в пределе, когда число бросаний стремится к бесконечности.*

Одно утверждение такого типа было установлено еще в рамках теории Лапласа; оно является следствием теоремы Муавра — Лапласа, о которой говорилось в § 9.

Пусть ε — некоторое положительное число, и пусть $p_n(\varepsilon)$ — вероятность того, что частота выпадения герба при n бросаниях отличается от 1/2 больше чем на ε. Тогда $p_n(\varepsilon)$ стремится к нулю, если n стремится к бесконечности:

$$\lim_{n\to\infty} p_n(\varepsilon)=0.$$

Эта теорема (доказанная впервые Я. Бернулли и известная под названием *закона больших чисел*) устанавливает следующее: каково бы ни было (положительное) ε, вероятность того, что отклонение частоты от вероятности превосходит ε, можно сделать сколь угодно малой, взяв достаточно большое число испытаний. Обоснование этого закона с точки зрения теории меры тривиально (по крайней мере до тех пор, пока речь идет о бросании монеты); поскольку интересующие нас события соответствуют конечным множествам.

Однако тут же возникает более сложный вопрос: какова вероятность того, что частота (в пределе при стремящемся к бесконечности числе испытаний) действительно равна $1/2$?

В качестве пространства выборок Ω мы вынуждены в этом случае рассматривать множество *всех бесконечных* слов вида *ГРГГР*...; нас интересует множество A, состоящее из тех слов, для которых

$$\lim_{n \to \infty} \frac{k_n(Г)}{n} = \frac{1}{2},$$

где $k_n(Г)$ — число букв $Г$ среди первых n букв слова.

Измеримо ли множество A (т. е. можно ли приписать ему меру, согласованную с заданием мер элементарных событий)? Наше требование счетной аддитивности обеспечивает положительный ответ на этот вопрос. Как можно в этом убедиться?

Для построения искомой меры нужна совокупность элементарных событий, вероятности (меры) которых известны или заданы; такую совокупность доставляет нам элементарная (и классическая) часть теории вероятностей.

Элементарные события соответствуют множествам слов, в которых фиксированы первые m букв ($m = 1, 2, \ldots$). Вероятность такого события в предположении, что монета честная, а бросания независимы, равна $1/2^m$. Достаточно ли «богата» множествами такая совокупность, чтобы дать нам возможность приписать меру множеству A?

Сопоставим букве $Г$ цифру 1, а букве P — цифру 0. Тогда мы можем «закодировать» каждое слово некоторой последовательностью нулей и единиц, скажем 10110...; такую последовательность в свою очередь можно рассматривать как запись в двоичной системе счисления некоторого действительного числа t, заключенного между 0 и 1. Таким способом можно установить соответствие между действительными числами t, такими, что $0 \leqslant t \leqslant 1$, и бесконечными словами из двух букв. Это соответствие можно сделать взаимно однозначным, если раз и навсегда условиться, какое из двух бесконечных двоичных представлений выбирать в тех случаях, когда такой вопрос возникает. Например,

$$1/4 = 0{,}01000\ \ldots \quad \text{и} \quad 1/4 = 0{,}00111\ \ldots,$$

и в этом случае необходимо сделать какой-то выбор.

Использование двоичной системы диктуется не только соображениями простоты[1]). Решающим здесь служит то обстоятель-

[1]) Воспользовавшись троичной системой, множество бесконечных слов из двух букв можно было бы отобразить в канторовское совершенное множество, которое получается из интервала (0, 1) выбрасыванием сначала его средней трети (исключая концы), затем средних третей (снова за исключением концов) двух оставшихся интервалов и т. д. до бесконечности.

ство, что, как легко проверить, элементарные события отображаются при этом в отрезки, длины которых равны соответствующим вероятностям. В самом деле, определенный выбор первых m букв слова соответствует заданию первых m двоичных цифр числа, а множеством действительных чисел, первые m цифр которых фиксированы, является интервал

$$(l/2^m,\ (l+1)/2^m)$$

(длины $1/2^m$), где l может равняться $0, 1, 2, \ldots, 2^m-1$ в зависимости от того, как именно выбраны цифры.

Таким образом, мера в пространстве выборок Ω всех бесконечных слов из двух букв переходит в обычную меру Лебега на интервале $(0, 1)$, а нахождение вероятности того, что

$$\lim_{n\to\infty} \frac{k_n(\Gamma)}{n} = \frac{1}{2},$$

эквивалентно нахождению меры Лебега множества тех чисел из интервала $(0, 1)$, которые в своем двоичном разложении имеют асимптотически одинаковое число нулей и единиц [1]).

В ответе получается 1, и можно сказать, что с вероятностью 1 частота выпадения герба в бесконечной последовательности независимых бросаний честной монеты равна $^1/_2$.

Эта теорема носит название *усиленного закона больших чисел*: ее открытие (в 1909 г. Эмилем Борелем) знаменует собой начало новой эры в развитии теории вероятностей.

Усиленный закон больших чисел постигла судьба всех великих открытий в математике: его существенно обобщили и расширили, что привело к новым задачам и стимулировало поиски новых методов. Это был первый серьезный и отважный шаг, выводящий из круга проблем, унаследованных от Лапласа, и этот шаг стал возможным лишь благодаря развитию теории меры. Последнее же в свою очередь стало возможным только потому, что математики приобщились к языку теории множеств.

Пространство всех бесконечных последовательностей из нулей и единиц *бесконечномерно* в том смысле, что для описания каждой его «точки» требуется бесконечно много «координат». Мы занимались построением в этом пространстве счетно аддитивной меры, которая была бы «естественной» с точки зрения задачи о независимых бросаниях честной монеты. Эта точка зрения немедленно подсказывает расширение на более общие бесконечномерные пространства, в которых каждая координата сама является элементом некоторого более общего, чем $\{0, 1\}$, множества и даже необязательно представляет собой число. Общая теория независимых испытаний может быть сформулирована в терминах так называемых *произведений мер* в таких бесконечномерных пространствах.

Но математическое воображение на этом не остановилось. Оно привело к мерам в множествах кривых. Наиболее известная

---

[1]) Мы уже упоминали об этой задаче в § 10.

и интересная из таких мер была введена в начале 20-х годов Норбертом Винером в связи с теорией броуновского движения.

С тех пор математики нашли новые неожиданные применения меры Винера в совсем, казалось бы, не связанных с ней областях своей науки. Например, выяснилось, что мера Винера множества кривых, исходящих из некоторой точки p пространства и достигающих в некоторый момент времени трехмерной области R, равна электростатическому потенциалу в точке p, порожденному таким распределением заряда, при котором граница «проводника» R является эквипотенциальной поверхностью с потенциалом 1. Но вычисление такого потенциала можно классическими методами свести к решению некоторого дифференциального уравнения; тем самым установлена связь между классическим анализом и теорией меры. В этом направлении уже многое сделано и еще больше делается сейчас, на пороге 70-х годов[1]).

§ 13. Группы и преобразования

Одним из самых важных, плодотворных и всеобъемлющих математических понятий является понятие *группы*. Это понятие удобно ввести в связи с понятием *преобразования* множества в себя.

Пусть S — какое-то множество; под преобразованием f множества S в себя понимается некоторый способ сопоставления каждому элементу p множества S *единственного* элемента $f(p)$ этого множества; $f(p)$ называется *образом элемента p при преобразовании f*.

Если $f(p) = p$ при *каждом* p, то f — *тождественное преобразование*. Если преобразование $f(p)$ взаимно однозначно, т. е. $f(p) \neq f(q)$ всякий раз, когда $p \neq q$, то можно определить *обратное преобразование* f^{-1}:

$$f^{-1}(q) = p, \quad \text{если} \quad q = f(p).$$

Иначе говоря, образом элемента q при преобразовании f^{-1} является тот (единственный) элемент, образом которого служит q при преобразовании f.

Если заданы два преобразования f и g множества S в себя, то можно определить новое преобразование fg, которое получится в результате последовательного выполнения сначала g, а затем f. Иначе говоря, обра-

[1]) Американское издание книги вышло в 1968 г. — *Прим. перев.*

зом элемента p при преобразовании fg служит образ при преобразовании f образа при преобразовании g этого элемента, т. е.

$$fg(p) = f(g(p)).$$

Кроме того, можно определить преобразование gf:

$$gf(p) = g(f(p)),$$

причем gf, вообще говоря, *не совпадает* с fg.

Таким образом, операция *композиции* преобразований (т. е. взятия fg или gf) *некоммутативна*, однако эта операция *ассоциативна*:

$$(fg)h = f(gh).$$

Это можно проверить, взяв какой-нибудь элемент p множества S и убедившись в том, что его образы при преобразованиях $(fg)h$ и $f(gh)$ совпадают.

Говорят, что (конечная или бесконечная) совокупность G преобразований образует *группу*, если выполняются следующие условия:

1. Вместе с любыми двумя преобразованиями f и g совокупность G содержит и их композицию fg (и, конечно, gf).

2. Тождественное преобразование принадлежит G.

3. Вместе с каждым преобразованием f совокупность G содержит и обратное к нему преобразование f^{-1}.

Если множество S конечно, то любое взаимно однозначное преобразование его в себя сводится к изменению *порядка* его элементов. Такие преобразования называются *подстановками*. Если S состоит из n объектов, то число различных подстановок равно $n! = 1 \cdot 2 \cdot 3 \ldots n$. Обозначая эти объекты числами $1, 2, \ldots, n$, будем записывать подстановку f в виде

$$f = \begin{pmatrix} 1 & 2 & 3 & \ldots & n \\ f(1) & f(2) & f(3) & \ldots & f(n) \end{pmatrix},$$

где $f(k)$ — образ k при подстановке f. Ясно, что если $i \neq j$, то $f(i) \neq f(j)$ (поскольку f — взаимно однозначное преобразование множества S в себя).

Вот, например, шесть $(3! = 1\cdot 2\cdot 3 = 6)$ подстановок множества из трех элементов:

$$f_0 = \begin{pmatrix} 1 & 2 & 3 \\ 1 & 2 & 3 \end{pmatrix}, \quad f_1 = \begin{pmatrix} 1 & 2 & 3 \\ 1 & 3 & 2 \end{pmatrix}, \quad f_2 = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & 3 \end{pmatrix},$$

$$f_3 = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix}, \quad f_4 = \begin{pmatrix} 1 & 2 & 3 \\ 3 & 1 & 2 \end{pmatrix}, \quad f_5 = \begin{pmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{pmatrix}.$$

Заметим, что f_0 — тождественное преобразование; кроме того, $f_1^2 = f_2^2 = f_5^2 = f_0$ (т. е. $f_1^{-1} = f_1$, $f_2^{-1} = f_2$, $f_5^{-1} = f_5$), $f_3 f_4 = f_0$, $f_1 f_4 = f_2$, $f_4 f_1 = f_5$ и т. д. Для шести подстановок $f_0, \ldots, f_5$ можно выписать такую таблицу умножеиия:

	f_0	f_1	f_2	f_3	f_4	f_5
f_0	f_0	f_1	f_2	f_3	f_4	f_5
f_1	f_1	f_0	f_4	f_5	f_2	f_3
f_2	f_2	f_3	f_0	f_1	f_5	f_4
f_3	f_3	f_2	f_5	f_4	f_0	f_1
f_4	f_4	f_5	f_1	f_0	f_3	f_2
f_5	f_5	f_4	f_3	f_2	f_1	f_0

Группы подстановок были впервые введены Абелем и Галуа в связи с их выдающимися работами по разрешимости алгебраических уравнений в радикалах. Некоторые идеи, лежащие в основе этих работ, настолько важны и настолько отвечают духу алгебры, что мы приведем здесь их краткий обзор.

Рассмотрим кубическое уравнение

$$x^3 + ax^2 + bx + c = 0.$$

Оно имеет три, вообще говоря, различных корня x_1, x_2, x_3. Коэффициенты a, b, c выражаются через эти корни по известным формулам

$$a = -(x_1 + x_2 + x_3),$$
$$b = x_1x_2 + x_2x_3 + x_3x_1,$$
$$c = -x_1x_2x_3.$$

Функция $F(x_1, x_2, x_3)$ называется *симметрической*, если ее значение не меняется при любой подстановке аргументов x_1, x_2, x_3. Таким образом, коэффициенты кубического уравнения являются симметрическими функциями его корней. Произвольная функция корней, вообще говоря, не будет симметрической. Естественно поставить вопрос: при каких подстановках она не меняется?

Например, функция

$$\Delta \equiv (x_1 - x_2)(x_1 - x_3)(x_2 - x_3)$$

не меняется при подстановках f_0, f_3 и f_4, а при остальных подстановках меняет знак.

Функция

$$\Phi \equiv a_1x_1 + a_2x_2 + a_3x_3$$

меняется, вообще говоря, при любых подстановках, кроме f_0.

Рассмотрим подстановки, оставляющие неизменной некоторую заданную функцию. Они образуют *подгруппу*, т. е. подмножество группы всех подстановок, которое само является группой (иначе говоря, подгруппа содержит тождественную подстановку f_0, вместе с каждыми f_i и f_j ей принадлежит f_if_j, и, кроме того, она содержит обратный к каждому из своих элементов).

Допустим теперь, что некоторый многочлен Ψ инвариантен относительно подгруппы (f_0, f_3, f_4), оставляющей неизменной функцию Δ. Мы докажем, что тогда Ψ имеет вид

$$\Psi = A(x_1, x_2, x_3) + B(x_1, x_2, x_3)\Delta,$$

где A и B — *симметрические* многочлены.

Наше предположение состоит в том, что для *всех* x_1, x_2, x_3

$$\Psi(x_1, x_2, x_3)=\Psi(x_2, x_3, x_1)=\Psi(x_3, x_1, x_2);$$

поэтому, заменяя x_1 на x_2, а x_2 на x_1, получаем

$$\Psi(x_2, x_1, x_3)=\Psi(x_1, x_3, x_2)=\Psi(x_3, x_2, x_1).$$

Кроме того, можно считать, что (вообще говоря)

$$\Psi(x_2, x_1, x_3)\neq\Psi(x_1, x_2, x_3),$$

ибо в противном случае функция Ψ была бы симметрической, а наше утверждение — тривиальным (доказываемое соотношение выполнялось бы при $B=0$).

Итак, многочлен $\Psi(x_1, x_2, x_3) - \Psi(x_2, x_1, x_3)$ не равен нулю тождественно, но принимает значение 0 при $x_2 = x_1$ и потому *делится* на $x_2 - x_1$. Аналогично можно показать, что он делится также на $x_1 - x_3$ и $x_2 - x_3$ и, следовательно, на Δ. Таким образом,

$$\Psi(x_1, x_2, x_3)-\Psi(x_2, x_1, x_3)=B(x_1, x_2, x_3)\Delta,$$

а поскольку

$$\Psi(x_2, x_1, x_3)=\Psi(x_1, x_3, x_2)=\Psi(x_3, x_2, x_1),$$

то и

$$\Psi(x_1, x_2, x_3)-\Psi(x_1, x_3, x_2)=B(x_1, x_2, x_3)\Delta,$$
$$\Psi(x_1, x_2, x_3)-\Psi(x_3, x_2, x_1)=B(x_1, x_2, x_3)\Delta.$$

Отсюда видно, что B не меняется при подстановках f_1, f_2 и f_5, а так как $f_4 = f_1 f_2$ и $f_3 = f_2 f_1$, то мы заключаем, что B не меняется при *всех* подстановках и, следовательно, есть симметрическая функция.

Аналогично доказывается, что $\Psi(x_1, x_2, x_3) - B\Delta$ — симметрическая функция; обозначая ее через A, мы получаем требуемое представление

$$\Psi = A + B\Delta.$$

Мы остановились столь подробно на выводе этого частного результата потому, что в основе этого вывода лежит необычайно мощная и плодотворная идея: можно многое узнать о *структуре* некоторых определенных объектов, исследуя только их *поведение под действием некоторой группы преобразований.*

Например, в физике изучение группы преобразований, относительно которых инвариантны силы, связывающие вместе атомы в молекуле, позволяет сделать далеко идущие выводы о поведении спектров молекул (скажем, объяснить так называемые правила отбора). В обширном новом мире элементарных частиц, даже не зная основных взаимодействий, можно получить глубокое представление о положении вещей, если постулировать симметрию относительно некоторой группы преобразований (вспомним в связи с этим о случае с группой SU_3, получившем широкую известность[1])).

Вернемся к алгебраическим уравнениям и покажем, как наши рассуждения приводят к доказательству разрешимости кубического уравнения в радикалах. Пусть

$$\omega = -\frac{1}{2} + \frac{\sqrt{3}}{2} i,$$

так что

$$\omega^2 = -\frac{1}{2} - \frac{\sqrt{3}}{2} i,$$

$$\omega^3 = 1.$$

Рассмотрим функцию $\Psi = (x_1 + \omega x_2 + \omega^2 x_3)^3$ и заметим, что она не меняется при подстановках f_3 и f_4. В самом деле, применяя к ней f_3, мы получаем $(x_2 + \omega x_3 + \omega^2 x_1)^3 = \omega^3 (x_2 + \omega x_3 + \omega^2 x_1)^3 = (\omega x_2 + \omega^2 x_3 + x_1)^3$. Следовательно, $(x_1 + \omega x_2 + \omega^2 x_3)^3 = A(x_1, x_2, x_3) + B(x_1, x_2, x_3)\Delta$, и можно показать, что A, B и Δ^2 выражаются в виде многочленов от a, b и c. Таким образом,

$$x_1 + \omega x_2 + \omega^2 x_3 = \sqrt[3]{A + B\Delta}\,.$$

---

[1]) Авторы имеют в виду теоретическое предсказание элементарной частицы омега-минус-барион, сделанное на основе математических соображений, связанных с изучением представлений группы SU_3, и подтвержденное затем открытием этой частицы. Подробнее об этом см., например, Ф. Дж. Дайсон, Математика в физических науках, в сб. «Математика в современном мире», изд-во «Мир», М., 1967, стр. 111—127. — *Прим. перев.*

Замечая, что функция $(x_1+\omega^2x_2+\omega x_3)^3$ тоже не меняется при подстановках f_3 и f_4, получаем

$$x_1+\omega^2x_2+\omega x_3=\sqrt[3]{\bar{A}+\bar{B}\Delta},$$

где $\bar{A}$ и $\bar{B}$ снова выражаются в виде многочленов от коэффициентов уравнения. Поскольку, как замечено выше,

$$x_1+x_2+x_3=-a,$$

можно разрешить эти уравнения относительно x_1, x_2, x_3 и получить формулы, содержащие кубические и квадратные корни. Следует помнить, что многочленом от коэффициентов является Δ^2, поэтому само Δ уже включает в себя квадратный корень.

Если проделать в явном виде перечисленные операции, получатся известные формулы Кардано, открытые в 16-м веке. В то время связь между алгебраическими уравнениями и группами преобразований была еще неизвестна. Установление этой связи Галуа и Абелем раскрыло движущие пружины сделанных задолго до них открытий и навело на мысль о возможности распространения этих методов на уравнения четвертой и высших степеней.

Всякую подстановку n объектов можно представить в виде произведения транспозиций (т. е. подстановок, меняющих местами два объекта и оставляющих остальные на месте). Хотя такое представление, вообще говоря, не единственно, одна и та же подстановка не может разлагаться и на четное, и на нечетное число транспозиций. Это приводит к естественному разделению подстановок на *четные* (состоящие из четного числа транспозиций) и *нечетные* (состоящие из нечетного числа транспозиций). Все четные подстановки (составляющие половину всей группы) образуют так называемую *знакопеременную* подгруппу. В случае $n=3$ это подстановки f_0, f_3, f_4. Функция, которая не меняется при подстановках из знакопеременной подгруппы, снова имеет вид $A+B\Delta$, где A и B — симметрические функции, а Δ есть произведение *всех* разностей $x_i - x_j$:

$$\Delta = (x_1 - x_2)(x_1 - x_3) \ldots (x_{n-1} - x_n).$$

Теперь можно попытаться найти функцию (подобную функции $x_1 + \omega x_2 + \omega^2 x_3$), какая-либо степень которой не меняется при подстановках из знакопеременной подгруппы. При $n = 4$ можно найти квадратичную функцию от x_1, x_2, x_3, x_4, куб которой обладает этим свойством. Однако при $n \geqslant 5$ *таких функций не существует*, и это является первым указанием на то, что уравнения степени выше четвертой неразрешимы в радикалах. Доказательство того, что не существует функции, степень которой не меняется при подстановках из знакопеременной подгруппы, опирается *только* на свойства группы подстановок более чем четырех объектов.

В качестве иллюстрации гибкости и многосторонности понятий подстановки и группы подстановок рассмотрим вкратце проблему классификации правил бракосочетания в первобытных обществах. В результате очень интересных исследований нескольких первобытных племен (выполненных, главным образом, Леви-Штраусом и его сотрудниками) был открыт следующий набор общих правил:

(1) Каждому члену племени приписывается определенный *брачный тип*, и только индивидуумам одного и того же типа разрешается вступать в брак.

(2) Тип индивидуума *однозначно* определяется его полом и типом его родителей.

(3) Если две родительские пары принадлежат к разным типам, то их сыновья (и соответственно дочери) тоже принадлежат к разным типам.

(4) Закон, разрешающий или запрещающий мужчине жениться на родственнице, зависит только от вида родства; в частности, мужчине не разрешается жениться на своей сестре.

(5) Для любых двух индивидуумов можно указать таких их потомков, которым разрешается вступать друг с другом в брак.

Ясно, что законы бракосочетания определяются, во-первых, тем, сколько брачных типов t_1, t_2, ..., t_n

используется в данном обществе, и, во-вторых, заданием правила, позволяющего определить тип $S(t)$ (соотв. $D(t)$) сына (соотв. дочери), если t — тип родителей.

Правила (2) и (3) показывают, что $S(t)$ и $D(t)$ являются подстановками, а из первой части правила (4) следует, что либо $S(t)$ $(D(t))$ совпадает с t при любом t (т. е. является *тождественной подстановкой*), либо $S(t)$ $(D(t))$ отличается от t при любом t (такая подстановка называется *полной*). Дальнейшее рассмотрение правил бракосочетания приводит к выводу, что S и D следует выбирать так, чтобы порождаемая ими группа состояла только из тождественной подстановки I и полных подстановок и была *транзитивной*, т. е. для любых t_i и t_j содержала такую подстановку P, что $P(t_i)=t_j$. Например, в обществе Тарау $S(t)\equiv t$, а $D(t)$ определяется так:

$$D=\begin{pmatrix} t_1 & t_2 & t_3 & t_4 \\ t_4 & t_1 & t_2 & t_3 \end{pmatrix}.$$

Заметим, что

$$D^2=\begin{pmatrix} t_1 & t_2 & t_3 & t_4 \\ t_3 & t_4 & t_1 & t_2 \end{pmatrix}, \quad D^3=\begin{pmatrix} t_1 & t_2 & t_3 & t_4 \\ t_2 & t_3 & t_4 & t_1 \end{pmatrix}, \quad D^4=I,$$

и, следовательно, группа состоит из I, D, D^2 и D^3. В другом первобытном обществе также используются четыре брачных типа, но S и D определяются следующим образом:

$$S=\begin{pmatrix} t_1 & t_2 & t_3 & t_4 \\ t_3 & t_4 & t_1 & t_2 \end{pmatrix}, \quad D=\begin{pmatrix} t_1 & t_2 & t_3 & t_4 \\ t_4 & t_3 & t_2 & t_1 \end{pmatrix}.$$

Изучение групп подстановок привело к построению общей и абстрактной теории групп. Абстрактной группой называется множество элементов $f_0, f_1, f_2, \ldots$ с заданной в нем бинарной операцией, сопоставляющей каждой паре элементов f_i и f_j однозначно определенный третий элемент $f_i \cdot f_j$, также являющийся элементом этой группы. При этом предполагается выполнение следующих условий:

1. $(f_i \cdot f_j) \cdot f_k = f_i \cdot (f_j \cdot f_k)$ (ассоциативность).

2. Существует единственный *нейтральный элемент* f_0, такой, что $f_0 \cdot f_i = f_i \cdot f_0 = f_i$ при любом i.

3. Для всякого элемента f_i имеется единственный элемент $\hat{f}_i$ (его иногда называют *обратным* к элементу f_i), такой, что

$$f_i \cdot \hat{f}_i = \hat{f}_i \cdot f_i = f_0.$$

Даже определенную абстрактно группу можно представлять себе как некоторую группу подстановок. Например, если элементы группы можно выписать в определенном порядке $f_0, f_1, f_2, \ldots$, то каждому f_i можно поставить в соответствие подстановку

$$\begin{pmatrix} f_0 & f_1 & f_2 & \ldots \\ f_i \cdot f_0 & f_i \cdot f_1 & f_i \cdot f_2 & \ldots \end{pmatrix}.$$

Нетрудно проверить, что подстановка, соответствующая элементу $f_i \cdot f_j$, является композицией подстановок, соответствующих элементам f_i и f_j (с учетом порядка), и, следовательно, исходная группа и отвечающая ей группа подстановок неразличимы (изоморфны).

Искусство математического доказательства часто состоит в нахождении подходящей схемы, в рамках которой доказываемый факт становится почти очевидным. Математическое творчество во многом заключается в поисках таких схем. Иногда их находят в богатом мире материальных объектов, иногда же (и это самая высшая форма творчества) их придумывают. Нередки и такие случаи, когда удается понять, что изучаемое явление укладывается в уже существующую схему, введенную раньше совсем для других целей. (Когда какая-нибудь схема повторяется в разных контекстах, она становится теорией и ее начинают изучать уже ради нее самой.)

Прекрасной иллюстрацией служит применение понятия группы в элементарной теории чисел. Рассмотрим, например, теорему Вильсона: если p — простое число, то $(p-1)!+1$ (т. е. $1 \cdot 2 \ldots (p-1)+1$) делится на p. (Если, скажем, $p=7$, то $(p-1)!+1 = 721$, а это число, очевидно, делится на 7.)

В формулировке этой теоремы нет никакого намека на связь с теорией групп. Однако существует простой способ сопоставить каждому простому числу некоторую группу — так называемую *группу вычетов по модулю p*. Ее элементами являются числа 1, 2, 3, ..., $p-1$, а бинарная операция $\circ$ определяется так:

$$i \circ j = \text{остаток от деления } i \cdot j \text{ на } p,$$

где $i \cdot j$ — обычное умножение чисел. Например, в случае $p = 7$ «таблица умножения» для этой группы выглядит так:

	1	2	3	4	5	6
1	1	2	3	4	5	6
2	2	4	6	1	3	5
3	3	6	2	5	1	4
4	4	1	5	2	6	3
5	5	3	1	6	4	2
6	6	5	4	3	2	1

Нетрудно проверить, что операция $\circ$ (умножение по модулю p) удовлетворяет всем необходимым условиям. Кроме того, эта группа *коммутативна*, т. е.

$$i \circ j = j \circ i.$$

Рассмотрим теперь выражение

$$(1 \circ 2 \circ 3 \circ \ldots \circ (p-1))^2 =$$

$$= \underbrace{1 \circ 2 \circ 3 \circ \ldots \circ (p-1)}_{\mathrm{I}} \circ \underbrace{1 \circ 2 \circ 3 \circ \ldots \circ (p-1)}_{\mathrm{II}}.$$

Ввиду коммутативности мы можем переписать это произведение в таком порядке, чтобы после каждого элемента из I стоял обратный к нему элемент из II. Тогда ясно, что

$$(1 \circ 2 \circ 3 \circ \ldots \circ (p-1))^2 = 1;$$

иначе говоря, квадрат элемента $1 \circ 2 \circ \ldots \circ (p-1)$ является нейтральным элементом. Но если k — такой элемент группы, что $k \circ k = 1$, то $k^2 - 1 = (k-1)(k+1)$ делится на p. Поскольку $1 \leqslant k \leqslant p-1$, отсюда следует, что либо $k = 1$, либо $k = p - 1$, т. е.

$$1 \circ 2 \circ 3 \circ \ldots \circ (p-1) = 1 \text{ или } 1 \circ 2 \circ 3 \circ \ldots \circ (p-1) = p-1.$$

Если в произведении $1 \circ 2 \circ 3 \circ \ldots \circ (p-1)$ мы снова умножим каждый элемент на обратный к нему, у нас останется только $p-1$ — единственный, кроме 1, элемент, обратный самому себе. Следовательно, $1 \circ 2 \circ 3 \circ \ldots \circ (p-1) = p-1$, а это и есть, в слегка замаскированном виде, утверждение теоремы Вильсона.

Стоило только придумать группу вычетов по модулю p, как она привела к другим фактам теории чисел. Например, имеется простая теорема, утверждающая, что в конечной группе (не обязательно коммутативной), состоящей из g элементов, каждый элемент, умноженный g раз сам на себя, дает нейтральный элемент:

$$\underbrace{f_i \cdot f_i \cdot \ldots \cdot f_i}_{g \text{ раз}} = f_0.$$

Применяя эту теорему к группе вычетов по модулю p, состоящей из $p-1$ элементов, мы получаем для любого k $(1 \leqslant k \leqslant p-1)$

$$\underbrace{k \circ k \circ \ldots \circ k}_{p-1 \text{ раз}} = 1,$$

т. е. $k^{p-1} - 1$ делится на p. Это известная теорема, высказанная Ферма в 1640 г. Понятие группы позволило установить ее тесную связь с более поздней теоремой Вильсона.

В геометрии и в физике мы снова встречаемся с понятием группы. Группу составляют преобразования, сохраняющие расстояния и углы (движения). Преобразования пространства-времени, оставляющие инвариантными «световые конусы»

$$(x - x_0)^2 + (y - y_0)^2 + (z - z_0)^2 = c^2 (t - t_0)^2,$$

образуют группу Лоренца специальной теории относительности.

Феликс Клейн в своей знаменитой Эрлангенской программе предложил рассматривать геометрию как изучение инвариантов определенных групп преобразований. Так, по Клейну, евклидова геометрия — это изучение инвариантов группы, состоящей из переносов, поворотов и тех же преобразований, сопровождаемых осевой симметрией; проективная геометрия — изучение инвариантов группы так называемых проективных преобразований и т. д.

Плодотворность такой точки зрения объясняется тем, что алгебраические свойства группы преобразований, оставляющих инвариантной некоторую математическую структуру, отражают многие свойства самой этой структуры[1].

Роль понятия группы в современной математике трудно переоценить. Во всей этой науке нет уголка, в котором так или иначе не ощущалось бы заметное влияние теоретико-групповых соображений. Несмотря на скромность исходных аксиом, группы принесли нам огромные математические богатства, и еще большие сокровища остаются пока скрытыми. Теория групп — один из наиболее активно изучаемых разделов математики, и не проходит, пожалуй, ни одного дня, не приносящего открытия какого-нибудь нового результата или нового приложения этой теории.

§ 14. Группы гомологий

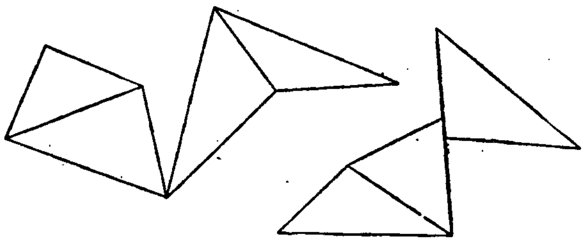
Совсем другую ситуацию, где понятие группы играет решающую роль, мы встречаем в топологии. Здесь мы можем дать лишь самое беглое элементар-

---

[1]) Преимущества такой точки зрения используют даже при исследовании абстрактных групп, когда изучают их автоморфизмы. *Автоморфизмом* группы G называется преобразование f этой группы в себя, сохраняющее групповую операцию, т. е. обладающее тем свойством, что $f(p \cdot q) = f(p) \cdot f(q)$ для любых $p, q \in G$. Автоморфизмы любой заданной группы G сами образуют группу, свойства которой отражают многие глубокие структурные свойства исходной группы G.

ное введение в важный и широко изученный раздел топологии, называемый *теорией гомологий*.

Основными геометрическими понятиями этой теории являются *симплексы* и *симплициальные комплексы*. Так, например, трехмерный симплекс — это тетраэдр, двумерный — треугольник; отрезок представляет собой одномерный симплекс, а точка — нульмерный симплекс. Симплициальным комплексом называется совокупность симплексов, такая, что любые два из них либо вообще не пересекаются, либо имеют общим



а *б*

Рис. 10.

целый симплекс меньшей размерности. Например, совокупность симплексов на рис. 10, *а* является симплициальным комплексом, а на рис. 10, *б* — не является, потому что там два треугольника имеют общей лишь часть стороны.

Симплекс называется *ориентированным*, если задано определенное упорядочение его вершин. При этом не все такие упорядочения считаются различными. Например, существует шесть различных способов ориентировать двумерный симплекс — по числу подстановок трех его вершин; однако упорядочения, которые можно получить одно из другого *четной* подстановкой (т. е. подстановкой, разложимой в произведение четного числа транспозиций), считаются эквивалентными. Для треугольника четными являются подстановки f_0, f_3 и f_4, и, следовательно, имеются только *две* различные ориентации. Очевидно, что нульмерные симплексы можно ориентировать лишь одним способом, однако нам удобно считать, что сами точки имеют две ориентации.

Если симплекс ориентирован (т. е. фиксировано некоторое упорядочение его вершин), то и каждый его подсимплекс автоматически становится ориентированным (индуцированная ориентация). Однако правило ориентации подсимплексов выглядит несколько неожиданно: пусть $p_1, p_2, \ldots$ — упорядочение вершин некоторого симплекса; тогда ориентация грани, противолежащей вершине p_k, задается тем же самым упорядочением остальных вершин, если k нечетно, и противоположна этой естественной ориентации, если k четно.

Это определение применяется к симплексам любых размерностей. Например, если для тетраэдра, изображенного на рис. 11, выбрана ориентация (1 2 3 4), то

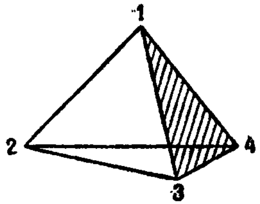


Рис. 11.

для заштрихованной грани следует брать не ориентацию (1 3 4), а ориентацию (1 4 3), которую мы будем записывать как —(1 3 4). Это правило и заставляет нас приписывать точке две ориентации. В самом деле, ориентация отрезка (одномерного симплекса) (1 2) индуцирует ориентацию (1) точки 1, а ориентация (2 1) индуцирует противоположную ориентацию —(1) этой точки.

Два ориентированных симплекса, пересекающиеся по некоторому симплексу, индуцируют на нем две ориентации, которые либо совпадают, либо противоположны.

Пусть $\sigma_1^{(k)}, \sigma_2^{(k)}, \ldots, \sigma_m^{(k)}$ — все ориентированные k-мерные симплексы некоторого симплициального комплекса K; k-мерной цепью называется формальное выражение вида

$$a_1\sigma_1^{(k)} + a_2\sigma_2^{(k)} + \ldots + a_m\sigma_m^{(k)},$$

где $a_1, a_2, \ldots, a_m$ — целые числа (положительные, отрицательные или нуль). Цепи одной и той же размерности можно «складывать», складывая коэффициенты при соответствующих симплексах, и относительно этой операции они образуют группу.

Для k-мерного ориентированного симплекса $\sigma_i^{(k)}$ его граница $\Delta\left(\sigma_i^{(k)}\right)$ определяется как цепь

$$\sigma_{i1}^{(k-1)} + \sigma_{i2}^{(k-1)} + \ldots + \sigma_{ir}^{(k-1)} = \Delta\left(\sigma_i^{(k)}\right),$$

где $\sigma_{i1}^{(k-1)}, \ldots, \sigma_{ir}^{(k-1)}$ — это $(k-1)$-мерные симплексы, составляющие геометрическую границу симплекса $\sigma_i^{(k)}$, взятые с ориентацией, индуцированной ориентацией $\sigma_i^{(k)}$. Например, границей тетраэдра, изображенного на рис. 11, является цепь

$$-(1\ 2\ 3) - (1\ 3\ 4) + (1\ 2\ 4) + (2\ 3\ 4).$$

Граница Δ цепи $a_1\sigma_1^{(k)} + \ldots + a_m\sigma_m^{(k)}$ определяется формулой

$$\Delta\left(a_1\sigma_1^{(k)} + \ldots + a_m\sigma_m^{(k)}\right) = a_1\Delta\left(\sigma_1^{(k)}\right) + \ldots + a_m\Delta\left(\sigma_m^{(k)}\right)$$

с учетом следующего правила: если $\sigma_i^{(k)}$ и $\sigma_j^{(k)}$ пересекаются по некоторому $(k-1)$-мерному симплексу, то он появится в этой «сумме» дважды — один раз с коэффициентом a_i и второй раз — с коэффициентом a_j; в случае, когда его индуцированные ориентации совпадают, следует брать коэффициент $a_i + a_j$, а когда они противоположны, $a_i - a_j$.

Вычислим, например, границу цепи $-(1\ 2\ 3) - (1\ 3\ 4) + (1\ 2\ 4) + (2\ 3\ 4)$, которая, как мы видели, сама является границей симплекса $(1\ 2\ 3\ 4)$. Мы имеем

$$\begin{aligned}
-\Delta(1\ 2\ 3) &= -(2\ 3) + (1\ 3) - (1\ 2),\\
-\Delta(1\ 3\ 4) &= -(3\ 4) + (1\ 4) - (1\ 3),\\
\Delta(1\ 2\ 4) &= (2\ 4) - (1\ 4) + (1\ 2),\\
\Delta(2\ 3\ 4) &= (3\ 4) - (2\ 4) + (2\ 3),
\end{aligned}$$

и, следовательно, искомая граница равна нулю. Это общее свойство цепей, являющихся границами; граница границы равна нулю, т. е.

$$\Delta\Delta = 0.$$

Рассмотрим теперь для некоторого симплициального комплекса совокупность B_r всех r-мерных цепей, являющихся границами $(r+1)$-мерных цепей, и совокупность Z_r всех r-мерных цепей, границы которых равны нулю (такие цепи называются циклами). Обе эти совокупности являются группами относительно сложения, причем группа циклов Z_r содержит группу B_r (поскольку граница границы равна нулю, и, следовательно, каждый элемент группы B_r содержится в множестве всех r-мерных цепей с нулевой границей).

Будем считать два цикла эквивалентными, если их разность принадлежит B_r (в частности, все элементы группы B_r считаются эквивалентными). Тогда множество Z_r разобьется на *непересекающиеся* классы эквивалентных циклов. (Это частный случай общего принципа «классов эквивалентности», который мы подробнее обсудим в гл. 2.) Сумма двух классов C_1 и C_2 определяется так: пусть c_1 — некоторый цикл из C_1, а c_2 — некоторый цикл из C_2; тогда $C_1 + C_2$ — это класс, содержащий $c_1 + c_2$. Классы эквивалентности образуют группу относительно этой новой операции сложения (нейтральным элементом этой группы служит класс B_r); это *факторгруппа* группы Z_r по B_r. Она обозначается

$$H^{(r)} = Z_r/B_r$$

и называется *r-мерной группой гомологий* симплициального комплекса.

Рассмотрим двумерный симплициальный комплекс, состоящий из четырех граней тетраэдра, изображенного на рис. 11, с такой ориентацией: (1 2 3), (1 2 4), (1 3 4), (2 3 4). Поскольку он не содержит трехмерных симплексов, группа B_2 состоит лишь из нейтрального элемента нуль[1]) и, следовательно,

$$H^{(2)} = Z_2,$$

где Z_2 — группа циклов, т. е. таких цепей

$$a(1\ 2\ 3) + b(1\ 2\ 4) + c(1\ 3\ 4) + d(2\ 3\ 4),$$

для которых

$$\Delta\{a(1\ 2\ 3) + b(1\ 2\ 4) + c(1\ 3\ 4) + d(2\ 3\ 4)\} = 0.$$

Как и выше, имеем

$$\begin{aligned}
&\Delta\{a(1\ 2\ 3) + b(1\ 2\ 4) + c(1\ 3\ 4) + d(2\ 3\ 4)\} = \\
&\quad = a(2\ 3) - a(1\ 3) + a(1\ 2) + b(2\ 4) - b(1\ 4) + b(1\ 2) + \\
&\qquad + c(3\ 4) - c(1\ 4) + c(1\ 3) + d(3\ 4) - d(2\ 4) + d(2\ 3) = \\
&\quad = (a+d)(2\ 3) + (-a+c)(1\ 3) + (a+b)(1\ 2) + \\
&\qquad + (b-d)(2\ 4) + (-b-c)(1\ 4) + (c+d)(3\ 4)
\end{aligned}$$

Для того чтобы это выражение обращалось в нуль, должны выполняться условия

$$\begin{aligned}
&a + d = 0, \qquad -a + c = 0, \qquad a + b = 0, \\
&b - d = 0, \qquad -b - c = 0, \qquad c + d = 0.
\end{aligned}$$

Эти уравнения не являются независимыми: уравнения во второй строке следуют из уравнений первой строки (например, складывая два первых уравнения первой строки, мы получаем последнее уравнение второй строки). Из первых трех уравнений находим

$$b = -a, \qquad c = a, \qquad d = -a.$$

Итак, двумерными циклами здесь являются цепи вида

$$a(1\ 2\ 3) - a(1\ 2\ 4) + a(1\ 3\ 4) - a(2\ 3\ 4).$$

Группа таких цепей относительно сложения *неотличима* от группы целых чисел и может быть отождествлена с ней, т. е.

$$H^{(2)} = \text{группа целых чисел по сложению.}$$

---

[1]) Можно считать, что B_2 — пустое множество, но удобнее рассматривать B_2 как группу, состоящую из границы цепи $0(1\ 2\ 3\ 4)$, т. е. элемента 0.

Более громоздкие вычисления позволяют установить, что $B_1 = Z_1 =$ группа целых чисел по сложению, и, таким образом,

$H^{(1)} = Z_1/B_1 =$ тривиальная группа, состоящая из единственного элемента 0.

Рассмотрим теперь симплициальный комплекс, состоящий только из трех ориентированных граней (1 2 3), (1 2 4), (1 3 4) тетраэдра, изображенного на рис. 11. Теперь обе группы $H^{(2)}$ и $H^{(1)}$ тривиальны: каждая из них состоит из одного нейтрального элемента.

С другой стороны, группы гомологий любого замкнутого выпуклого многогранника, гранями которого служат треугольники, совпадают с группами гомологий тетраэдра.

Оказывается, группы гомологий характеризуют *внутреннее* устройство комплекса, т. е. то, каким способом он «собран» из симплексов. Этим и объясняется их важная роль в топологии.

В топологии две геометрические конфигурации рассматриваются как тождественные, если между ними можно установить *непрерывное* взаимно однозначное соответствие. Такое соответствие называется *гомеоморфизмом*, и можно сказать, что в топологии отождествляются те конфигурации, которые гомеоморфны. Например, с топологической точки зрения тетраэдр и сфера тождественны. Если некоторая конфигурация допускает сколь угодно «хорошую» аппроксимацию симплициальными комплексами, то, как можно показать, все аппроксимирующие комплексы имеют одни и те же группы гомологий во всех размерностях; следовательно, можно говорить о группах гомологий таких конфигураций.

Важная теорема (восходящая к Пуанкаре) утверждает, что гомеоморфные конфигурации имеют одинаковые соответственные группы гомологий во всех размерностях.

Проиллюстрируем сказанное одним совсем простым примером. Рассмотрим две плоские кривые: простую замкнутую кривую, изображенную на рис. 12, и кривую с одной точкой самопересечения (восьмерку), изображенную на рис. 13. Ясно, что они не гомеоморфны (в самом деле, если удалить любую точку первой кривой, она останется связной, у второй же кривой есть точка, а именно точка самопересечения, отбросив которую мы нарушим ее связность). Посмотрим, как отражает теория гомологий это различие.

Простую замкнутую кривую мы можем приблизить *простым замкнутым* многоугольником. Такой

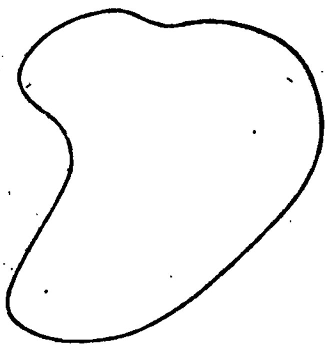


Рис. 12.

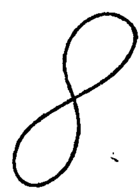


Рис. 13.

многоугольник представляет собой одномерный симплициальный комплекс (рис. 14). Его группой гомологий $H^{(1)}$ всегда является группа целых чисел по сложению, *независимо от числа сторон.* Таким образом, одномерная группа гомологий простой замкнутой кривой есть группа целых чисел по сложению.

Кривую в форме восьмерки уже нельзя приблизить простым многоугольником. Для такой кривой симплициальной аппроксимацией служит комплекс, изображенный на рис. 15. Для него одномерная группа гомологий есть группа *упорядоченных пар целых чисел* (a, b) относительно операции сложения, определенной обычным образом, а именно

$$(a_1, b_1) + (a_2, b_2) = (a_1 + a_2, b_1 + b_2).$$

Эта группа отличается от группы целых чисел, и, следовательно, простая замкнутая кривая не гомеоморфна восьмерке.

В то время как гомеоморфные конфигурации имеют одинаковые группы гомологий, обратное, вообще говоря, *не верно* для конфигураций размерности больше 2. Общая проблема определения, гомеоморфны или нет заданные конфигурации высших размерностей, остается нерешенной.

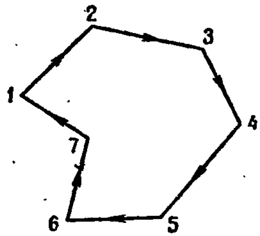


Рис. 14.

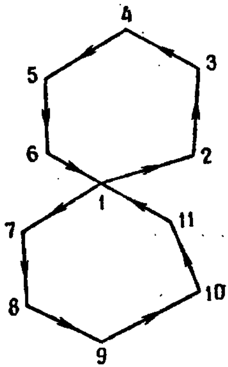


Рис. 15.

В этом параграфе мы рассмотрели важный пример тенденции к *алгебраизации математики.* Она состоит в том, что для решения задач геометрии и анализа подбирают и изучают подходящие алгебраические структуры (например, группы), дискретные и комбинаторные по своей природе. Эта тенденция весьма характерна для современной математики; мы вернемся к этому в гл. 2, где обсудим еще два примера.

§ 15. Векторы, матрицы и геометрия

Важной и весьма характерной для математики тенденцией является взаимное проникновение и последующая унификация ее на вид совершенно различных частей. Хорошим примером тому служит аналитическая геометрия, появившаяся благодаря Де-карту, который ввел в геометрию координаты. Введение координат позволило, с одной стороны, описывать геометрические объекты (такие, как конические сечения) при помощи алгебраических уравнений, а с другой стороны, интерпретировать алгебраические уравнения геометрически (скажем, рассматривать уравнение с двумя переменными как задающее некоторую кривую). Эта связь между алгеброй и геометрией, оказавшаяся особенно плодотворной в последнее время, и будет темой настоящего параграфа. Например, мы увидим, как можно интерпретировать геометрически задачу решения системы линейных уравнений. Центральную роль здесь играет абстрактное и весьма элегантное по своей простоте понятие линейного векторного пространства и линейные преобразования таких пространств. Важным конкретным представлением линейных преобразований служат так называемые матрицы. Эти математические объекты (мы определим и обсудим их в настоящем параграфе) являются одновременно и неразделимо как алгебраическими, так и геометрическими и потому объединяют две эти дисциплины.

Как мы уже подчеркивали, важным идеям свойственно находить применение в неожиданных областях; так обстоит дело и с идеями линейной алгебры, к которой относятся перечисленные выше понятия. В последующих параграфах мы увидим, как они применяются в специальной теории относительности и в теории марковских цепей — важном разделе современной теории вероятностей.

При дальнейшем изложении нам несколько раз представится случай упомянуть пространства размерности больше, чем 3; поэтому стоит ненадолго остановиться и подробнее обсудить эти объекты и некоторые тесно связанные с ними вопросы.

Введем прежде всего понятие *линейного векторного пространства*. Для начала рассмотрим обычную плоскость. Выберем на ней раз и навсегда некоторую точку O (рис. 16). С каждой точкой P плоскости мы связываем *направленный* отрезок (*вектор*), ведущий от O к P. Таким способом точки отождествляются с векторами. Введем теперь в множестве векторов две важные операции.

1. Умножение вектора на действительное число (скаляр) α определяется так: αP есть вектор, длина

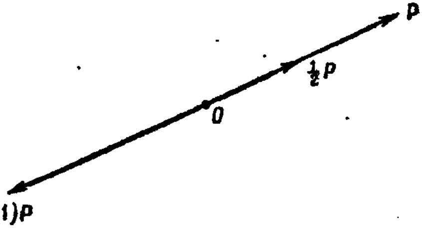


Рис. 16.

которого в $|\alpha|$ раз превосходит длину вектора P, причем αP принадлежит той же прямой, что и P. (Напомним, что $|\alpha|$ есть абсолютная величина числа α; скажем, $|-2| = 2$, $|2| = 2$.) Этот новый вектор направлен в ту же сторону, что и P, если $\alpha > 0$, и в противоположную сторону, если $\alpha < 0$.

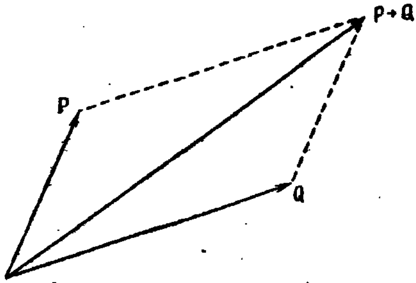


Рис. 17.

2. Сложение векторов определяется по обычному правилу параллелограмма (рис. 17).

Можно проверить, что эти две основные операции обладают следующими свойствами:

(a) $P + Q = Q + P$;

(b) $(P + Q) + R = P + (Q + R)$;

(c) существует единственный вектор 0, такой, что для каждого вектора P имеет место равенство $P + 0 = P$;
(d) для всякого вектора P существует единственный вектор P', такой, что $P' + P = 0$;
(e) $\alpha(P + Q) = \alpha P + \alpha Q$;
(f) $(\alpha + \beta)P = \alpha P + \beta P$;
(g) $\alpha(\beta P) = (\alpha\beta)P$;
(h) $(1)P = P$.

Теперь (как это часто делается в математике) мы можем обратить эту процедуру и определить линейное векторное пространство как множество объектов $P, Q, \ldots$, в котором определены операции умножения на скаляр и сложения, удовлетворяющие перечисленным восьми аксиомам.

Хотя мы пришли к этим аксиомам, рассматривая векторы на плоскости, оказалось, что им удовлетворяют и терхмерные векторы, а также объекты, мало похожие на векторы. Например, линейное векторное пространство образуют многочлены с действительными коэффициентами, непрерывные действительные функции, заданные на отрезке, а также цепи, определенные в § 14.

Тот факт, что такие разные объекты, как векторы на плоскости (или в пространстве) и непрерывные функции, образуют линейные векторные пространства, свидетельствует о том, что аксиомы (a) — (h) определяют структуру весьма общего вида. В дальнейшем мы увидим, что эту структуру можно обогатить, вводя дополнительные аксиомы.

Покажем теперь, как можно ввести понятие размерности. Для этого сначала определим понятие *линейной независимости*. Векторы $P_1, P_2, \ldots, P_n$ называются линейно независимыми, если из любого линейного соотношения

$$\alpha_1 P_1 + \alpha_2 P_2 + \ldots + \alpha_n P_n = 0$$

следует, что $\alpha_1 = \alpha_2 = \ldots = \alpha_n = 0$. Иными словами, ни один из векторов P_i не может быть представлен в виде линейного выражения (линейной комбинации), содержащего только остальные векторы рассматриваемой совокупности.

Линейное векторное пространство называется n-мерным, если в нем существуют n линейно независимых векторов, но *не существует совокупности $n+1$ линейно независимых векторов.*

Если линейное векторное пространство n-мерно, то n линейно независимых векторов P_1, P_2, ..., P_n называют его *базисом.* Это означает, что любой вектор P рассматриваемого пространства может быть единственным образом представлен в виде линейной комбинации векторов P_1, P_2, ..., P_n:

$$P = \alpha_1 P_1 + \alpha_2 P_2 + \ldots + \alpha_n P_n.$$

Если выбран некоторый базис, то каждому вектору P однозначно соответствует упорядоченный набор n скаляров $(\alpha_1, \alpha_2, \ldots, \alpha_n)$, и обратно. Базис называют также системой координат, а числа α_1, α_2, ..., α_n — координатами вектора P.

Линейное преобразование T векторного пространства в себя — это отображение, которое каждому вектору P ставит в соответствие вектор $T(P)$ того же пространства, причем так, что выполняются следующие условия:

$$T(P+Q) = T(P) + T(Q),$$
$$T(\alpha P) = \alpha T(P).$$

Используя базис (систему координат), каждому линейному преобразованию можно однозначно сопоставить квадратную таблицу чисел, называемую *матрицей.* Покажем, как это сделать. Выберем некоторый базис P_1, P_2, ..., P_n и рассмотрим «преобразованные» векторы $T(P_1)$, $T(P_2)$, ..., $T(P_n)$. Поскольку каждый из них является линейной комбинацией векторов базиса, мы можем написать:

$$\begin{aligned} T(P_1) &= t_{11}P_1 + t_{12}P_2 + \ldots + t_{1n}P_n, \\ T(P_2) &= t_{21}P_1 + t_{22}P_2 + \ldots + t_{2n}P_n, \\ &\cdots\cdots\cdots\cdots\cdots\cdots \\ T(P_n) &= t_{n1}P_1 + t_{n2}P_2 + \ldots + t_{nn}P_n. \end{aligned}$$

Таким образом, линейному преобразованию T соответствует (однозначно, если фиксирован некоторый базис) матрица чисел

$$T=\begin{pmatrix} t_{11} & t_{12} & \dots & t_{1n} \\ t_{21} & t_{22} & \dots & t_{2n} \\ \cdot & \cdot & \cdot & \cdot \\ t_{n1} & t_{n2} & \dots & t_{nn} \end{pmatrix}.$$

Пусть S — другое линейное преобразование рассматриваемого пространства, и пусть матрица преобразования S относительно выбранного базиса $P_1, P_2, \dots$ $\dots, P_n$ имеет вид

$$S=\begin{pmatrix} s_{11} & s_{12} & \dots & s_{1n} \\ s_{21} & s_{22} & \dots & s_{2n} \\ \cdot & \cdot & \cdot & \cdot \\ s_{n1} & s_{n2} & \dots & s_{nn} \end{pmatrix}.$$

Предположим, что выполняется сначала преобразование T, а затем S. Объединенное действие двух этих преобразований — их композиция ST — снова, как легко видеть, является линейным преобразованием. Преобразованию ST соответствует (относительно того же базиса $P_1, P_2, \dots, P_n$) матрица, элемент которой с индексом (i, k) (т. е. стоящий на пересечении i-й строки и k-го столбца) выражается формулой

$$s_{i1}t_{1k}+s_{i2}t_{2k}+\dots+s_{in}t_{nk}=\sum_{j=1}^{n} s_{ij}t_{jk}.$$

Если взять композицию преобразований S и T в обратном порядке (т. е. сначала выполнить S, а затем T), то (i, k)-й элемент матрицы, соответствующей TS, будет иметь вид

$$t_{i1}s_{1k}+t_{i2}s_{2k}+\dots+t_{in}s_{nk}=\sum_{j=1}^{n} t_{ij}s_{jk};$$

это выражение, вообще говоря, отличается от предыдущего $\sum\limits_{j=1}^{n} s_{ij}t_{jk}$.

Таким образом, *композиция* линейных преобразований соответствует определенной в множестве матриц операции, называемой *умножением* матриц.

Тождественное преобразование I представляется (в любом базисе) так называемой единичной или тождественной, матрицей

$$\begin{pmatrix} 1 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \\ . & . & . & . & . \\ 0 & 0 & \dots & 0 & 1 \end{pmatrix}.$$

Можно попытаться определить преобразование, обратное к T, как такое преобразование, которое в композиции с T дает тождественное преобразование I. Однако не для всякого линейного преобразования T существует обратное. Преобразования, не имеющие обратного, характеризуются тем свойством, что они аннулируют некоторый ненулевой вектор (т. е. образом этого вектора при преобразовании служит нулевой вектор). Иными словами, линейное преобразование T имеет обратное T^{-1} (и притом единственное) тогда и только тогда, когда из равенства $T(P)=0$ следует, что P — нулевой вектор. Такие преобразования называются невырожденными. Композиция двух невырожденных преобразований есть снова невырожденное преобразование; невырожденные преобразования составляют группу.

Обсудим подробнее случай действительного двумерного линейного векторного пространства. Выбрав раз и навсегда базис, мы сможем отождествить векторы с упорядоченными парами (α_1, α_2) действительных чисел, а линейные преобразования — с (2×2)-матрицами

$$\begin{pmatrix} t_{11} & t_{12} \\ t_{21} & t_{22} \end{pmatrix},$$

где $t_{11}, \dots, t_{22}$ — действительные числа. Утверждение, что преобразование T имеет обратное, равносильно утверждению, что система линейных уравнений

$$t_{11}\alpha_1 + t_{12}\alpha_2 = 0,$$
$$t_{21}\alpha_1 + t_{22}\alpha_2 = 0$$

имеет единственное (тривиальное) решение $\alpha_1 = \alpha_2 = 0$. Из линейной алгебры известно, что последнее имеет место тогда и только тогда, когда

$$t_{11}t_{22} - t_{21}t_{12} \neq 0.$$

Число $t_{11}t_{22} - t_{21}t_{12}$ называется *определителем* (или *детерминантом*) матрицы T и обозначается так:

$$t_{11}t_{22} - t_{12}t_{21} = \det T = \begin{vmatrix} t_{11} & t_{12} \\ t_{21} & t_{22} \end{vmatrix}.$$

Можно показать, что обратная матрица T^{-1} имеет вид

$$T^{-1} = \begin{pmatrix} \frac{t_{22}}{\det T} & -\frac{t_{12}}{\det T} \\ -\frac{t_{21}}{\det T} & \frac{t_{11}}{\det T} \end{pmatrix}$$

и, таким образом,

$$T^{-1}T = TT^{-1} = I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

Можно ли распространить понятие определителя на n-мерный случай и обобщить формулу обратной матрицы на матрицы порядка n? Исходным толчком к изучению такого рода вопросов послужило желание научиться решать системы n линейных уравнений с n неизвестными $\alpha_1, \alpha_2, \ldots, \alpha_n$ [1]:

$$\begin{aligned} t_{11}\alpha_1 + t_{12}\alpha_2 + \ldots + t_{1n}\alpha_n &= \beta_1, \\ t_{21}\alpha_1 + t_{22}\alpha_2 + \ldots + t_{2n}\alpha_n &= \beta_2, \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ t_{n1}\alpha_1 + t_{n2}\alpha_2 + \ldots + t_{nn}\alpha_n &= \beta_n. \end{aligned}$$

[1]) Мы надеемся, что читатель не будет возмущен нашим отказом от традиционного обозначения «неизвестных» величин буквами x, y и z. Хотя часто такие обозначения удобны, они ни к чему не обязывают; в конце концов обозначения — это только обозначения!

Для $n=2$, 3 и даже 4 неизвестные α_1, α_2, ... можно записать как отношения определенных алгебраических форм, содержащих t_{ij} и β_i, на вид довольно громоздких. Можно догадаться, а затем и доказать, что

$$\alpha_k = \frac{\det T_k}{\det T},$$

где матрица T_k получается из T заменой k-го столбца столбцом $(\beta_1, \beta_2, \ldots, \beta_n)$. При этом определитель общей $(n \times n)$-матрицы

$$A = \begin{pmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ \cdot & \cdot & \cdot & \cdot \\ a_{n1} & a_{n2} & \ldots & a_{nn} \end{pmatrix}$$

вычисляется по следующему правилу. Пусть

$$\pi = \begin{pmatrix} 1 & 2 & 3 & \ldots & n \\ i_1 & i_2 & i_3 & \ldots & i_n \end{pmatrix}$$

— некоторая подстановка индексов 1, 2, ..., n. Выпишем произведение

$$\pm a_{1i_1} a_{2i_2} a_{3i_3} \ldots a_{ni_n}$$

и возьмем его со знаком $+$, если подстановка π четна, и со знаком — в противном случае.

(Как говорилось выше, всякая подстановка представляется в виде произведения *транспозиций*, т. е. подстановок, меняющих местами два элемента и не изменяющих остальных. Хотя такое представление данной подстановки не единственно, оно всегда содержит либо только четное, либо только нечетное число транспозиций. В первом случае подстановка называется четной, во втором — нечетной.)

Определитель равен сумме всех таких произведений, взятых с соответствующими знаками, причем суммирование производится по всем $n!$ подстановкам:

$$\det A = \sum_{\pi} (\operatorname{sgn} \pi)\, a_{1\pi(1)} a_{2\pi(2)} \ldots a_{n\pi(n)},$$

где $\operatorname{sgn}\pi=+1$, если π четна, и $\operatorname{sgn}\pi=-1$, если π нечетна, а $\pi(1)$, $\pi(2)$, ... обозначают $i_1, i_2, \ldots$.

Введенный таким способом определитель выступает как некий алгебраический монстр, созданный главным образом для того, чтобы решать системы линейных уравнений. (При вычислении определителей редко пользуются непосредственно определением. Имеется обширная литература, посвященная свойствам определителей и методам их вычисления.) Однако в теории линейных векторных пространств определители можно рассматривать с более естественной геометрической точки зрения.

В самом деле, пусть $P_1, P_2, \ldots, P_n$ — некоторый базис нашего векторного пространства, а T — линейное преобразование этого пространства в себя. Пусть $Q_1, Q_2, \ldots, Q_n$ — другой базис. Интересно узнать, как меняются представления векторов и матриц при переходе от одного базиса к другому. Поскольку векторы P_j образуют базис, векторы Q_i выражаются в виде линейных комбинаций P_j:

$$Q_i=d_{i1}P_1+d_{i2}P_2+\ldots+d_{in}P_n=\sum_{j=1}^{n}d_{ij}P_j;$$

поэтому переход от одного базиса к другому можно описать матрицей

$$D=\begin{pmatrix} d_{11} & d_{12} & \ldots & d_{1n} \\ d_{21} & d_{22} & \ldots & d_{2n} \\ \cdot & \cdot & \cdot & \cdot \\ d_{n1} & d_{n2} & \ldots & d_{nn} \end{pmatrix}.$$

С другой стороны, поскольку Q_i тоже образуют базис, мы имеем

$$P_i=\sum_{j=1}^{n}c_{ij}Q_j,$$

где числа c_{ij} составляют матрицу

$$C=\begin{pmatrix} c_{11} & c_{12} & \ldots & c_{1n} \\ c_{21} & c_{22} & \ldots & c_{2n} \\ \cdot & \cdot & \cdot & \cdot \\ c_{n1} & c_{n2} & \ldots & c_{nn} \end{pmatrix}.$$

Можно проверить, что

$$CD = DC = I,$$

откуда $D = C^{-1}$ и $C = D^{-1}$, т. е. матрицы C и D взаимно обратны. Итак, n линейных комбинаций базисных векторов сами образуют базис тогда и только тогда, когда составленная из их коэффициентов матрица D невырожденна, т. е. когда существует обратная ей матрица D^{-1}, или, что равносильно, когда $\det D \neq 0$.

Пусть теперь

$$T = \begin{pmatrix} t_{11} & t_{12} & \dots & t_{1n} \\ t_{21} & t_{22} & \dots & t_{2n} \\ \cdot & \cdot & \cdot & \cdot \\ t_{n1} & t_{n2} & \dots & t_{nn} \end{pmatrix}$$

— матрица, описывающая линейное преобразование T в базисе $P_1, P_2, \dots, P_n$; тогда можно показать, что матрица

$$C^{-1}TC = DTD^{-1}$$

описывает *то же* преобразование T, но уже в базисе $Q_1, \dots, Q_n$.

Алгебраически матрицы $C^{-1}TC$ (где C пробегает все $(n \times n)$-матрицы, имеющие обратные; такие матрицы называются невырожденными) могут внешне очень сильно отличаться одна от другой. Но поскольку *известно*, что *геометрически* они описывают *одно и то же* линейное преобразование T, естественно думать, что они должны иметь нечто общее. В частности, возникает вопрос: нельзя ли построить из элементов матрицы такие выражения, которые не изменяются (инвариантны) при умножении этой матрицы справа на C и слева на C^{-1}. Оказывается, определитель является как раз таким выражением:

$$\det(C^{-1}TC) = \det T,$$

и, следовательно, он связан с преобразованием *внутренним* образом. Но тогда должен существовать способ, позволяющий ввести понятие определителя линейного преобразования, не обращаясь к системе координат (базису). Такой способ действительно существует и состоит в следующем.

Сначала пытаются найти (нулевой) вектор E, который при преобразовании T переходит в λE, где λ — скаляр. Иначе говоря, ищут векторные решения уравнения $TE = \lambda E$, $E \neq 0$. Можно показать, что в n-мерном векторном пространстве существуют n линейно независимых векторов $E_1, \ldots, E_n$ такого вида. Они называются *собственными векторами* преобразования T, а соответствующие скаляры $\lambda_1, \ldots, \lambda_n$ называются *собственными значениями*.

Ясно, что в системе координат $E_1, E_2, \ldots, E_n$ преобразование T описывается матрицей особенно простого вида, а именно — диагональной матрицей

$$\begin{vmatrix} \lambda_1 & 0 & 0 & \ldots & 0 \\ 0 & \lambda_2 & 0 & \ldots & 0 \\ 0 & 0 & \lambda_3 & \ldots & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & \ldots & \lambda_n \end{vmatrix}.$$

Ее определитель равен произведению диагональных элементов:

$$\det T = \lambda_1 \cdot \lambda_2 \ldots \lambda_n;$$

это выражение и можно считать алгебраическим определением определителя. Это *число*, отвечающее матрице, отражает некоторые важнейшие свойства описываемого ею преобразования. Например, линейное преобразование T имеет обратное тогда и только тогда, когда $\det T \neq 0$.

Дальше в этом параграфе мы еще вернемся к определителям, но сначала мы должны наделить наше линейное векторное пространство еще одной структурой.

До сих пор мы не употребляли такие слова, как *расстояние* или *угол*, исключая лишь то место, где определялось произведение αP. (Но даже и там без этого можно было обойтись. Мы могли бы определить nP для натуральных n как сумму $P + P + \ldots + P$

(n раз), а $\frac{1}{n}P$ — как вектор Q, который при сложении с самим собой n раз дает P. Таким способом можно было бы определить $\frac{m}{n}P$ для натуральных m и n. Располагая после этого определением rP для положительных рациональных r, мы могли бы определить $-rP$ как единственный вектор, который при сложении с rP дает 0. Чтобы распространить затем это определение на действительные числа, нужно было бы воспользоваться процедурой, кратко описанной в § 8, где мы говорили о расширении системы рациональных чисел до системы всех действительных чисел.)

Вспомиим, как обращаются с понятиями расстояния и угла в аналитической геометрии на плоскости. Сначала выбирают две перпендикулярные прямые (оси x и y) и каждой точке P приписывают координаты x и y (рис. 18). Затем, используя теорему Пи-

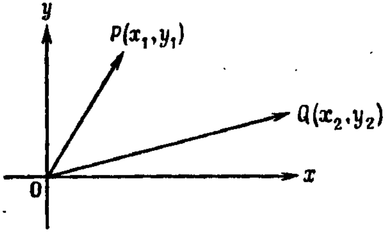


Рис. 18.

фагора, доказывают, что расстояние между точками $P(x_1, y_1)$ и $Q(x_2, y_2)$ задается формулой

$$d(P, Q) = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}.$$

Применяя правило косинусов (небольшое обобщение теоремы Пифагора), показывают, что угол θ между OP и OQ можно найти по формуле

$$\cos\theta = \frac{x_1x_2 + y_1y_2}{\sqrt{x_1^2 + y_1^2}\,\sqrt{x_2^2 + y_2^2}}.$$

Косинус угла между отрезками AP и AQ, где A — точка с координатами x_0, y_0 (рис. 19), находится по несколько более общей формуле

$$\cos\theta = \frac{(x_1-x_0)(x_2-x_0)+(y_1-y_0)(y_2-y_0)}{\sqrt{(x_1-x_0)^2+(y_1-y_0)^2}\sqrt{(x_2-x_0)^2+(y_2-y_0)^2}}.$$

Теперь можно попытаться обратить обычную процедуру и рассматривать эти формулы как определения расстояния и угла. При этом возникают две проблемы.

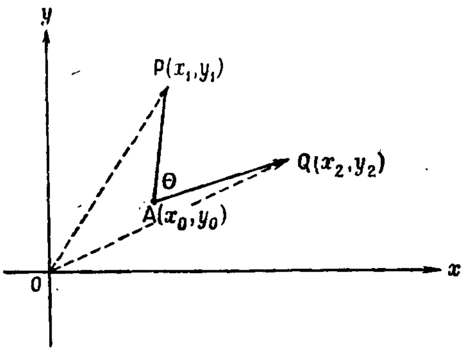


Рис. 19.

1. Поскольку эти формулы отнесены к некоторой выбранной системе координат, следует проверить, что они не меняются при переходе к другой системе координат того же типа. Иначе говоря, взяв другую систему (рис. 20) взаимно перпендикулярных осей x' и y', в которой точка P имеет координаты x'_1, y'_1, Q — координаты x'_2, y'_2 и 0 — координаты x'_0, y'_0, мы должны получить

$$(x'_2-x'_1)^2+(y'_2-y'_1)^2=(x_2-x_1)^2+(y_2-y_1)^2;$$

аналогичное тождество должно получиться из формулы для $\cos\theta$.

2. Как определяется *конгруэнтность* геометрических фигур — понятие, играющее центральную роль во всем построении геометрии?

Эти две проблемы тесно связаны, и решение второй из них почти автоматически приводит к решению первой.

В основе понятия конгруэнтности лежит интуитивное представление о *жестких движениях*, т. е. движениях, не изменяющих *метрические* соотношения в геометрических фигурах. Точнее, жесткие движения — это преобразования, сохраняющие расстояния и углы.

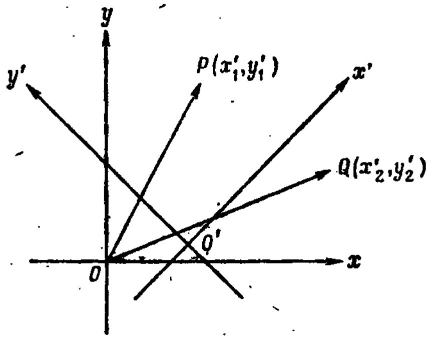


Рис. 20.

Можно показать, что всякое преобразование, сохраняющее расстояния и косинусы углов, имеет вид

$$x' = a_{11}x + a_{12}y + x_0,$$
$$y' = a_{21}x + a_{22}y + y_0,$$

где матрица

$$A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$$

такова, что $A'A = I$; через A' здесь обозначена транспонированная матрица

$$A' = \begin{pmatrix} a_{11} & a_{21} \\ a_{12} & a_{22} \end{pmatrix},$$

т. е. матрица, полученная из A отражением относительно главной диагонали (идущей из левого верхнего угла в правый нижний).

Можно проверить, что $\det A$ равен либо $+1$, либо -1. Если $\det A = -1$, то преобразование A (хотя оно и сохраняет расстояния и косинусы углов) не описывает жесткое движение плоскости. Так, например, преобразование

$$x' = x,$$
$$y' = -y$$

есть симметрия относительно оси x, сохраняющая d и $\cos\theta$, но меняющая ориентацию на противоположную (вращение против часовой стрелки переходит во вращение по часовой стрелке и наоборот). Вообще, любое такое преобразование A с $\det A = -1$ обращает ориентацию (рис. 21).

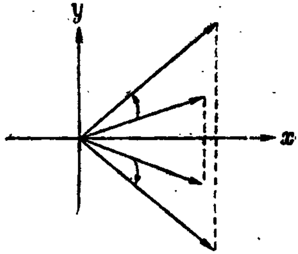


Рис. 21.

Итак, жесткими движениями являются преобразования вида

$$x' = a_{11}x + a_{12}y + x_0,$$
$$y' = a_{21}x + a_{22}y + y_0,$$

для которых $A'A = I$ и $\det A = 1$.

Оказывается, что переход к другой (прямоугольной) системе координат описывается формулами того же общего вида, что и жесткие движения, и, таким образом, проблема 1 решена.

Теперь, когда у нас есть *явное* и *конкретное* определение конгруэнтности, мы должны связать его с евклидовым понятием конгруэнтности.

Евклид не определяет конгруэнтность и не говорит нам, что это такое. Вместо этого он лишь перечисляет (в форме аксиом) свойства, которыми она *должна обладать*, чтобы соответствовать интуиции.

На основе этих и других аксиом (включая знаменитый пятый постулат, утверждающий, что через точку, не лежащую на прямой l, можно провести одну и только одну прямую, параллельную l) можно было бы доказать теорему Пифагора, опираясь только на понятие *конгруэнтности*. Одно такое доказательство демонстрируется на рис. 22; квадрат, построенный на

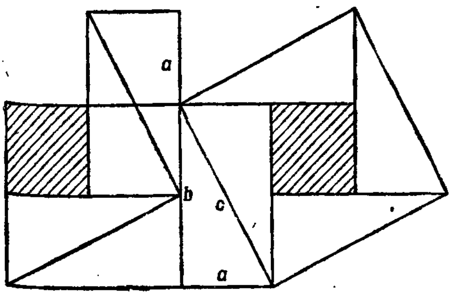


Рис. 22.

гипотенузе прямоугольного треугольника, и фигуру, составленную из квадратов, построенных на катетах (меньший из них приложен сверху к большему), можно разбить на пять взаимно конгруэнтных частей (четыре из них представляют собой копии исходного треугольника, а пятая — некоторый квадрат).

Чтобы получить из этого *нечислового* варианта теоремы Пифагора известный *числовой* вариант $c^2 = a^2 + b^2$, понадобится *теория измерения*, которую тоже можно построить на основе аксиом, не включающих чисел как таковых. Именно здесь у греков возникли трудности с иррациональными числами, поскольку аксиомы, которыми они пользовались, позволяли строить только рациональные числа.

Тот факт (вытекающий из теоремы Пифагора), что гипотенуза равнобедренного прямоугольного треугольника несоизмерима с его катетом (иррациональность числа $\sqrt{2}$), явился для греков источником глубокого огорчения. Сейчас мы знаем, что для построения полной теории измерения, допускающей наряду с рациональными и иррациональные числа, нужна новая весьма тонкая аксиома — так называемая *аксиома непрерывности*. Евклид и греки пропустили еще целую группу аксиом, называемых аксиомами порядка. Они относятся к таким понятиям, как «точка лежит между двумя другими точками», «луч OA лежит между двумя другими лучами OP и OQ» и т. д. Выражаемые ими свойства настолько интуитивно очевидны, что их просто не заметили, и это неудивительно. Однако вычислительная машина неспособна «увидеть», лежит ли данная точка между двумя другими, и ей должно быть «сказано», как обращаться с понятием «лежит между».

Здесь снова необязательно знать, что означает «лежать между», если имеется достаточно полный перечень свойств этого понятия.

Итак, следует проверить, что наша алгебраически определенная конгруэнтность обладает всеми свойствами, постулированными Евклидом. Коль скоро это сделано и коль скоро проверены все другие аксиомы, мы получаем полную алгебраическую (или аналитическую) *модель* плоской евклидовой геометрии.

Теперь уже легко шагнуть за пределы двух измерений, и лучше всего сделать это в рамках теории линейных векторных пространств. Единственное новое понятие, которое для этого потребуется, — это понятие *скалярного произведения* $\omega(P, Q)$ двух векторов P и Q. Оно вводится при помощи аксиомы, утверждающей существование величины $\omega(P, Q)$, обладающей такими свойствами:

(a) $\omega(P, Q)$ есть симметрическая билинейная функция от P и Q, т. е. $\omega(P, Q) = \omega(Q, P)$ и

$$\omega(P, \alpha Q + \beta R) = \alpha\omega(P, Q) + \beta\omega(P, R);$$

(b) $\omega(P, P) \geqslant 0$, причем $\omega(P, P) = 0$ только тогда, когда $P = 0$.

Покажем, что двумерное векторное пространство, в котором определено скалярное произведение

$\omega(P, Q)$, обладающее указанными свойствами, можно превратить в модель евклидовой плоскости.

Рассмотрим два линейно независимых вектора P_1 и P_2. Как нетрудно доказать, можно найти такие линейные комбинации E_1 и E_2 векторов P_1 и P_2, что

$$\omega(E_1, E_1) = \omega(E_2, E_2) = 1,$$
$$\omega(E_1, E_2) = 0.$$

Для доказательства нужно воспользоваться тем, что для линейно независимых векторов P_1 и P_2 имеет место неравенство

$$\omega(P_1, P_1)\,\omega(P_2, P_2) - \omega^2(P_1, P_2) > 0.$$

Это одна из форм известного неравенства Шварца.

Найденные так E_1 и E_2 тоже образуют базис; поэтому

$$P = x_1E_1 + y_1E_2,$$
$$Q = x_2E_1 + y_2E_2,$$

и мы получаем

$$\omega(P, Q) = x_1x_2 + y_1y_2.$$

Определим теперь квадрат расстояния между векторами P и Q формулой

$$d^2(P, Q) = \omega(P - Q, P - Q),$$

а косинус угла между P и Q — формулой

$$\cos\theta = \frac{\omega(P, Q)}{\sqrt{\omega(P, P)}\sqrt{\omega(Q, Q)}};$$

мы вернулись к обычным формулам плоской аналитической геометрии.

Теперь мы можем определить n-мерное евклидово пространство как *n-мерное действительное векторное пространство со скалярным произведением* $\omega(P, Q)$, *обладающим указанными выше свойствами* (a) *и* (b).

Жесткие движения — это *переносы* и *линейные преобразования* T (по аналогии называемые вращениями), сохраняющие скалярное произведение

$$\omega(TP, TQ) = \omega(P, Q)$$

и ориентацию, т. е. такие, что $\det T = 1$. (То, что $\det T = \pm 1$, следует из равенства $\omega(TP, TQ) = \omega(P,Q)$.)

Вращения и переносы образуют группу, и, следуя Эрлангенской программе Феликса Клейна, можно сказать, что евклидова геометрия (в пространстве любой размерности) есть изучение тех свойств геометрических фигур, которые инвариантны относительно этой группы.

Переносы и все невырожденные линейные преобразования (т. е. преобразования, имеющие единственное обратное) также образуют группу, гораздо более обширную, чем евклидова. Свойств, инвариантных относительно этой группы, меньше; они являются предметом так называемой *аффинной геометрии* (соответствующая группа называется аффинной группой).

В аффинной геометрии не существует способа различать окружность и эллипс (или два эллипса между собой), поскольку окружность можно перевести в эллипс аффинным преобразованием, а законным предметом аффинной геометрии являются только те свойства, которые при таких преобразованиях не меняются. Однако гипербола и эллипс различаются: эти кривые не переводятся одна в другую аффинным преобразованием.

Проективная геометрия еще более «примитивна». Поскольку она изучает свойства, инвариантные относительно проектирований, в ней не различаются даже эллипс и гипербола: обе эти кривые являются коническими сечениями, и, следовательно, их можно получить одну из другой при помощи проектирования.

Возвращаясь к n-мерной евклидовой геометрии, сделаем в заключение несколько замечаний.

Пусть $E_1, E_2, \ldots, E_n$ — ортонормированный базис, т. е.

$$\omega(E_i, E_i) = 1, \qquad i = 1, 2, \ldots, n,$$
$$\omega(E_i, E_j) = 0, \qquad i \neq j.$$

Иными словами, ортонормированный базис есть множество n *взаимно ортогональных* (перпендикулярных) *единичных* векторов. Существование такого базиса должно быть и может быть доказано. Элементы такого базиса линейно независимы.

Единичный куб, построенный на $E_1, E_2, \ldots, E_n$, есть множество векторов

$$P = x_1E_1 + \ldots + x_nE_n,$$

таких, что

$$0 < x_i < 1, \qquad i = 1, 2, \ldots, n.$$

«Прямоугольный параллелепипед», построенный на $E_1, E_2, \ldots, E_n$, — это множество векторов $P = x_1E_1 + \ldots + x_nE_n$, таких, что

$$0 < x_1 < a_1, \quad 0 < x_2 < a_2, \ldots, 0 < x_n < a_n.$$

Объем такого параллелепипеда по определению равен $a_1a_2 \ldots a_n$ (в частности, объем единичного куба равен 1).

Приписав объемы «прямоугольным параллелепипедам», мы можем при помощи аксиомы аддитивности (и аксиомы дополнения) распространить понятие объема (или, точнее, n-мерной меры Лебега) на широкую совокупность множеств. На самом деле, как можно показать, выбор величины $a_1a_2 \ldots a_n$ в качестве объема прямоугольного параллелепипеда — это единственный возможный выбор, совместимый с (a) аксиомой аддитивности, (b) требованием, чтобы конгруэнтные множества имели одинаковые меры, и (c) условием, чтобы объем изменялся непрерывно при изменении длины стороны.

В результате линейного пребразования T нашего единичного куба мы получим параллелепипед, который, вообще говоря, будет уже «косым» (т. е. не прямоугольным). Его объем оказывается равным $\pm \det T$, где знак $+$ или $-$ берется так, чтобы все выражение было положительным.

Вообще, в результате линейного преобразования T некоторого множества Ω получается множество $T(\Omega)$, мера которого равна произведению меры Ω на $\det T$ с подходящим знаком:

$$m(T(\Omega)) = \pm \det T \cdot m(\Omega).$$

Таким образом, странная алгебраическая форма

$$\sum_{\pi} \operatorname{sgn} \pi \, t_{1\pi(1)} t_{2\pi(2)} \ldots t_{n\pi(n)},$$

к которой приводит решение систем линейных уравнений, выступает в новом, в высшей степени привлекательном, геометрическом свете.

Многие свойства определителей, которые выводятся с большим трудом при помощи разных манипуляций с матрицами, становятся почти очевидными, как только они получили геометрическое истолкование.

Упомянем в качестве примера теорему (которую мы в этом параграфе несколько раз неявно использовали) о том, что определитель произведения двух матриц равен произведению их определителей:

$$\det(TS) = \det T \cdot \det S.$$

Алгебраическое доказательство трудоемко и затемняет смысл теоремы. Геометрически же это утверждение очевидно: оно устанавливает, что искажение объема, вызванное последовательным выполнением двух линейных преобразований, равно произведению искажений при каждом из них. Однако не следует думать, что можно получить что-нибудь «задаром». Теорема об определителе произведения становится очевидной только *после* того, как доказана теорема об искажении:

$$m(T(\Omega)) = \pm \det T \cdot m(\Omega),$$

а эта теорема не является ни очевидной, ни слишком простой!

Против геометрического доказательства можно выдвинуть тот довод, что прежде чем сделать его вразумительным, приходится вводить слишком много постороннего материала. В самом деле, к чему все эти n-мерные параллелепипеды и их объемы, если теорему можно доказать элементарными (пусть и несколько утомительными) алгебраическими выкладками?

Ну что ж, на такие вопросы нельзя ответить. Логически доказательство есть доказательство, и справедливость теоремы не зависит от того, каким способом она доказана (при условии, конечно, что строго соблюдаются логические правила игры).

Но к счастью, как мы неоднократно подчеркивали, математика — это больше, чем чистая логика. «Аромат» теоремы во многом зависит от контекста, в котором она формулируется, даже если это не отражается на ее истинности. Именно контекст позволяет отличить проблему от головоломки, а стройную теорию — от собрания случайных фактов.

§ 16. Специальная теория относительности как пример геометрического подхода в физике

Покажем теперь, как, модифицируя и обобщая некоторые понятия, рассмотренные в предыдущем параграфе, можно математически построить специальную теорию относительности.

Специальная теория относительности Эйнштейна возникла из желания примирить ньютоновскую механику с волновой теорией света.

Коротко говоря, дилемма состояла в следующем. В ньютоновской механике все наблюдатели, движущиеся один относительно другого прямолинейно и равномерно, равноправны. Если у каждого из этих наблюдателей имеется блокнот, в котором он записывает результаты своих наблюдений и измерений *механических явлений*, и если эти записи позднее сравниваются, они оказываются во всех отношениях идентичными.

С другой стороны, в волновой теории света требовалось наличие среды, в которой свет мог бы распространяться, и существование такой среды, названной «светоносным эфиром», специально постулировалось. С эфиром связывалась предпочтительная система отсчета, и, казалось бы, должен был существовать способ, позволяющий обнаружить при помощи световых сигналов равномерное движение наблюдателя относительно эфира. В частности, можно было надеяться обнаружить движение Земли сквозь эфир, сравнивая время, затраченное световым сигналом на то, чтобы пройти путь l в направлении этого движения, а затем обратно, и такой же путь в двух взаимно противоположиых направлениях, перпендикулярных движению Земли. Хотя эта разница во времени выражается весьма малой величиной порядка $(v/c)^2$ (где v — скорость Земли, а c — скорость света), ее можно было бы измерить при помощи хорошего интерферометра.

Такой эксперимент был проведен в 1887 г. Майкельсоном и Морли и привел к отрицательному результату! Отрицательный исход эксперимента Майкельсона — Морли поверг физику в кризис, который, по-видимому, был полиостью разрешен лишь в 1905 г., когда Эйнштейи сформулировал специальную теорию относительности.

Эйнштейн предложил сохранить принцип равиоправности наблюдателей, движущихся один относительио другого прямолинейно и равномерно (принцип относительности). Но он предложил также рассматривать результат эксперимента Майкельсона — Морли как новый закон, утверждающий, что скорость распространения света в вакууме (c) для всех этих наблюдателей одииакова.

Чтобы объединить в гармоничное целое прииципп относительности и прииципп постоянства скорости света, потребовался глубокий пересмотр представлеиий о пространстве и времени. Уловить, в чем тут дело, легче всего, воспользовавшись геометрическим подходом, предложениым Германом Мииковским.

Пусть S — прямоугольная система координат, а S' — другая такая система, движущаяся отиосительно первой прямолинейио и равномерно с постояииой скоростью v в положительном иаправлеиии оси x (рис. 23). Пусть, далее, O — наблюдатель, неподвижный относительио системы S, а O' — наблюдатель, неподвижиый относительно S'. Каждый из этих иаблюдателей имеет свой эталон длины. Представим себе, что в каждой из этих систем имеется множество синхронно идущих часов, расположенных настолько плотно, иасколько это потребуется.

Всякому событию наблюдатель O может сопоставить (упорядоченную) четверку действительных чисел (x, y, z, t), первые три из которых указывают, *где* произошло это событие, а последнее — *когда* оно произошло. То же событие для наблюдателя O' выразится другой четверкой чисел (x', y', z', t'), полученных при помощи *его* эталона длины и *его* часов.

Возникает вопрос: как связаны между собой две эти четверки чисел?

Если встать на ньютоновскую точку зрения и считать время абсолютным, то часы в этих двух системах

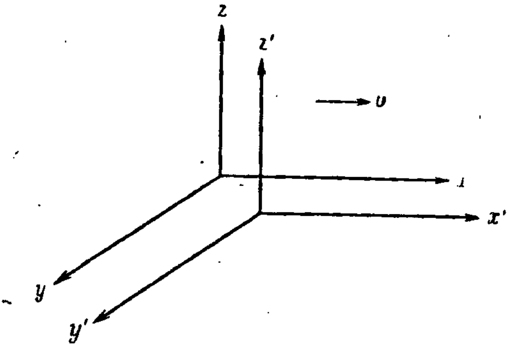


Рис. 23.

можно сделать синхронными и тогда

$$t' = t.$$

Кроме того, мы, очевидно, получим

$$y' = y,$$
$$z' = z,$$
$$x' = x - vt.$$

Такое преобразование координат x, y, z, t в координаты x', y', z', t' называется преобразованием Галилея; все законы классической динамики инвариантны относительно этого преобразования.

Однако, если наблюдатель O произвел вспышку света в точке $(0, 0, 0, 0)$ (которая по нашему предположению является общей для двух рассматриваемых систем), то фронт волны движется в пространстве по закону

$$x^2 + y^2 + z^2 = c^2t^2.$$

Наблюдая распространение этого светового сигнала, O' должен, согласно Эйнштейну, получить то же самое уравнение в координатах со штрихами:

$$x'^2 + y'^2 + z'^2 = c^2t'^2,$$

но это *несовместимо с преобразованием Галилея.*

Попытаемся найти такое преобразование координат x, y, z, t в x', y', z', t', при котором

$$x'^2 + y'^2 + z'^2 = c^2t'^2$$

всякий раз, когда

$$x^2 + y^2 + z^2 = c^2t^2.$$

Прежде всего, по аналогии с преобразованием Галилея, допустим, что это новое преобразование тоже линейно[1]). Затем можно удостовериться, что на самом деле имеет место тождество

$$x'^2 + y'^2 + z'^2 - c^2t'^2 = x^2 + y^2 + z^2 - c^2t^2$$

и что, как и раньше,

$$y' = y, \qquad z' = z.$$

Наконец, можно показать, что

$$x' = a_{11}x + a_{12}t,$$

$$t' = a_{21}x + a_{22}t$$

(т. е. в выражения для x' и t' не входят y и z). Далее, для *всех* x и t

$$x'^2 - c^2t'^2 = (a_{11}x + a_{12}t)^2 - c^2(a_{21}x + a_{22}t)^2 = x^2 - c^2t^2,$$

и, следовательно,

$$a_{11}^2 - c^2a_{21}^2 = 1,$$

$$a_{12}^2 - c^2a_{22}^2 = -c^2,$$

$$a_{11}a_{12} - c^2a_{21}a_{22} = 0.$$

Наблюдатель в системе S, видя, как движется начало системы S' ($x' = 0, y' = 0, z' = 0$), находит, что $x = vt$, т. е. из равенства $x' = 0$ *вытекает* равенство $x - vt = 0$.

[1]) Его линейность можно обосновать многими способами, но мы не будем здесь этим заниматься.

Поскольку x' есть линейная комбинация x и t (т. е. $x' = a_{11}x + a_{12}t$), отсюда следует, что

$$x' = \gamma(x - vt),$$

где γ может зависеть (и на самом деле зависит) от v. Тогда

$$a_{11} = \gamma, \quad a_{12} = -\gamma v,$$

откуда при помощи простых выкладок получаем:

$$a_{21} = \pm \frac{1}{c}\sqrt{\gamma^2 - 1},$$

$$a_{22} = \mp \frac{\gamma^2 v}{c}\sqrt{\frac{1}{\gamma^2 - 1}},$$

$$\gamma^2 = \frac{1}{1 - (v^2/c^2)}.$$

Для t' находим выражение

$$t' = \pm \frac{1}{c}\sqrt{\gamma^2 - 1}\, x \mp \frac{\gamma^2 v}{c}\sqrt{\frac{1}{\gamma^2 - 1}}\, t,$$

где знак выбирается так, чтобы при убывании величины v/c получалась формула, все более близкая к формуле $t' = t$ преобразования Галилея. Наконец, после ряда стандартных упрощений, получаем

$$x' = \frac{1}{\sqrt{1 - (v/c)^2}}(x - vt),$$

$$t' = \frac{1}{\sqrt{1 - (v/c)^2}}\left(t - \frac{v}{c^2}x\right).$$

Это — не что иное, как простейшая форма известных преобразований Лоренца. Для того чтобы x' и t' были действительными числами, должно выполняться неравенство $v < c$. В этом и состоит один из главных принципов теории относительности: относительная скорость наблюдателя не может превышать скорость света.

Одно из наиболее поразительных следствий формул Лоренца — относительность одновременности. События, которые наблюдатель O зарегистрировал как одновременные (одно и то же t и разные x), не покажутся такими наблюдателю O'!

Тесно связаны с этим и другие, не менее удивительные следствия — лоренцево сокращение эталонов длины и «растяжение времени».

Допустим, что наблюдатель O' отметил на своей оси x две точки $(x_1', 0, 0)$ и $(x_2', 0, 0)$ и нашел расстояние между ними:

$$l = x_2' - x_1'.$$

Наблюдатель O, пытаясь измерить это расстояние, мог бы попросить своих помощников зарегистрировать в некоторый *определенный момент времени* t (помощники, разумеется, имеют в своем распоряжении *синхронные* часы) координаты x (в системе S) точек, отмеченных наблюдателем O'. Помощники сообщат ему результаты:

$$x_1 = \sqrt{1-(v/c)^2}\, x_1' + vt,$$
$$x_2 = \sqrt{1-(v/c)^2}\, x_2' + vt,$$

и наблюдатель O, вычислив расстояние между точками, получит

$$x_2 - x_1 = \sqrt{1-(v/c)^2}\,(x_2' - x_1') = \sqrt{1-(v/c)^2}\, l,$$

т. е. величину, меньшую, чем найденная наблюдателем O'!

Аналогично, оказывается, что движущиеся часы идут медленнее (снова в $1/\sqrt{1-(v/c)^2}$ раз), чем часы, неподвижные относительно наблюдателя.

Фундаментальная важность преобразований Лоренца покоится на принципе, утверждающем, что *все законы физики должны быть инвариантны относительно этих преобразований*[1]. Хотя это просто другая

---

[1]) Для простоты мы рассматриваем здесь только весьма частный случай преобразований Лоренца, а именно преобразования, связывающие системы с общей осью x, когда скорость направлена вдоль этой общей оси. Они образуют только небольшую подгруппу так называемой однородной группы Лоренца, состоящей из всех линейных преобразований, оставляющих инвариантной форму $x^2 + y^2 + z^2 - c^2t^2$, *без* дополнительного требования, что $y' = y$ и $z' = z$. Подразумевается, конечно, что законы физики должны быть инвариантны относительно этой большей группы.

формулировка принципа относительности, она устанавливает глубокую аналогию между физикой и геометрией.

Как евклидова геометрия есть изучение группы линейных преобразований, не изменяющих евклидовы расстояния и углы, так и физика есть изучение инвариантов группы Лоренца, относительно которой инвариантна форма $x^2+y^2+z^2-c^2t^2$.

Аналогия становится еще более поразительной, если сравнить плоскую геометрию и двумерное пространство-время (x, t).

В плоской евклидовой геометрии вращения системы координат описываются такими матрицами

$$A=\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix},$$

что $A'A=I$, где A' — транспонированная матрица A. Можно показать, что все такие матрицы имеют вид

$$A=A(\theta)=\begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix},$$

где θ можно отождествить с углом, на который поворачивается система координат.

Матрица, определяемая преобразованием Лоренца (в данном частном случае), имеет вид

$$\begin{bmatrix} \dfrac{1}{\sqrt{1-(v/c)^2}} & -\dfrac{v}{\sqrt{1-(v/c)^2}} \\ \dfrac{-v}{c^2\sqrt{1-(v/c)^2}} & \dfrac{1}{\sqrt{1-(v/c)^2}} \end{bmatrix}.$$

Если вместо x, t и x', t' взять x, ict и x', ict', то преобразование примет вид

$$x'=\frac{1}{\sqrt{1-(v/c)^2}}\left(x+\frac{iv}{c}ict\right),$$

$$ict'=\frac{1}{\sqrt{1-(v/c)^2}}\left(-\frac{iv}{c}x+ict\right)$$

и ему будет соответствовать матрица

$$\begin{pmatrix} \dfrac{1}{\sqrt{1-(v/c)^2}} & \dfrac{iv}{c}\dfrac{1}{\sqrt{1-(v/c)^2}} \\ -\dfrac{iv}{c}\dfrac{1}{\sqrt{1-(v/c)^2}} & \dfrac{1}{\sqrt{1-(v/c)^2}} \end{pmatrix}.$$

Но можно найти такое действительное θ, что

$$\cos i\theta = \operatorname{ch}\theta = \frac{1}{\sqrt{1-(v/c)^2}},$$

$$\sin i\theta = -i\operatorname{sh}\theta = \frac{iv}{c}\frac{1}{\sqrt{1-(v/c)^2}}.$$

Тогда матрица примет вид

$$\begin{pmatrix} \cos i\theta & \sin i\theta \\ -\sin i\theta & \cos i\theta \end{pmatrix}$$

и соответствующее преобразование можно представлять себе как вращение на «мнимый» угол «системы координат (x, ict)». (Мы должны предостеречь читателя, что все это — просто терминология, которой не следует приписывать никакого мистического смысла, — ведь она введена единственно для того, чтобы подчеркнуть аналогию между евклидовой геометрией и теорией относительности.)

Чтобы понять, какую помощь оказывает обращение к геометрии при формулировке физических законов, рассмотрим вкратце явление упругого столкновения. Пусть две материальные точки масс m и M движутся (в системе S) вдоль оси x с постоянными скоростями u и U так, что происходит упругое столкновение.

Чтобы определить скорости u_1 и U_1 после столкновения, в классической динамике привлекают два закона:

(а) закон сохранения импульса

$$mu + MU = mu_1 + MU_1;$$

(b) закон сохранения энергии

$$\frac{1}{2}mu^2+\frac{1}{2}MU^2=\frac{1}{2}mu_1^2+\frac{1}{2}MU_1^2.$$

И до столкновения, и после него здесь существенна лишь кинетическая энергия, поэтому соотношение (b) выражает закон сохранения полной энергии. Полагая $\mu=m/M$, мы элементарными алгебраическими выкладками выводим, что решениями уравнений (a) и (b) являются либо $u_1=u$, $U_1=U$, либо

$$u_1=\frac{\mu-1}{\mu+1}u+\frac{2}{\mu+1}U,\quad U_1=\frac{2\mu}{\mu+1}u-\frac{\mu-1}{\mu+1}U.$$

Первое решение соответствует случаю, когда частицы вообще не сталкиваются, поэтому нас интересует только второе.

Понятие массы — очень тонкое понятие. На самом деле существуют три различных понятия: масса используется

(a) как мера «количества вещества»,

(b) как мера «сопротивления изменениям состояния движения»,

(c) как гравитационный «заряд».

В связи с (a) говорят о *собственной массе*, с (b) — об *инерциальной массе*, с (c) — о *гравитационной массе*.

В классической динамике эти три понятия не различают. Мы сейчас будем просто считать, что имеется некоторый способ приписывать «куску» вещества некоторое число (в некоторых единицах), называемое его *собственной массой*, и что в приведенные выше законы сохранения импульса и энергии входят именно эти собственные массы. Кроме того, мы предположим, что собственная масса сохраняется во всех физических процессах.

Теперь мы можем убедиться в том, что ни сохранение импульса, ни сохранение энергии в том виде, в каком они сформулированы выше, не являются законами физики. В самом деле, наблюдатель O', находящийся в системе S', рассматривая описанный выше процесс столкновения, найдет, что до столкновения скорости были равны

$$u' = \frac{u - v}{1 - \frac{uv}{c^2}}, \quad U' = \frac{U - v}{1 - \frac{Uv}{c^2}},$$

в то время как после столкновения они стали

$$u'_1 = \frac{u_1 - v}{1 - \frac{uv}{c^2}}, \quad U'_1 = \frac{U_1 - v}{1 - \frac{U_1 v}{c^2}}.$$

Чтобы было понятно, как получается, например, формула для u', заметим, что в системе S скорость u равна $\Delta x/\Delta t$ (смещению Δx, деленному на время Δt, за которое произошло это смещение). В системе S' мы имеем $u' = \Delta x'/\Delta t'$, где $\Delta x'$ и $\Delta t'$ получены из Δx и Δt преобразованием Лоренца. Следовательно,

$$u' = \frac{\Delta x - v\,\Delta t}{-\frac{v}{c^2}\Delta x + \Delta t} = \frac{u - v}{1 - \frac{uv}{c^2}}.$$

Теперь уже нетрудно проверить, что

$$mu' + MU' \neq mu'_1 + MU'_1,$$

$$\frac{1}{2} mu'^2 + \frac{1}{2} MU'^2 \neq \frac{1}{2} mu'^2_1 + \frac{1}{2} MU'^2_1.$$

Посмотрим, как идея инвариантности приводит к подходящей модификации закона сохранения импульса.

В классической физике скорость движущейся частицы есть векторная величина, описываемая в прямоугольной системе координат x, y, z тремя компонентами u_x, u_y, u_z.

С точки зрения теории относительности миру присуща четырехмерность, где роль четвертого измерения играет время. С другой стороны, пространственные и временные переменные связаны между собой преобразованиями Лоренца. Поэтому векторные величины, рассматриваемые в теории относительности, должны быть «4-векторами», т. е. объектами, которые в каждой системе отсчета описываются четырьмя компонентами

$$w_x, \quad w_y, \quad w_z, \quad w_t,$$

причем эти компоненты при переходе к другой системе отсчета преобразуются так же, как координаты x, y, z, t, т. е. по формулам Лоренца. Иначе говоря, в системе S' компоненты вектора w имеют вид

$$w_{x'} = \frac{w_x - v w_t}{\sqrt{1-(v/c)^2}},$$

$$w_{y'} = w_y, \quad w_{z'} = w_z,$$

$$w_{t'} = \frac{-(v/c^2)\, w_x + w_t}{\sqrt{1-(v/c)^2}}.$$

Теперь должно быть понятно, что не существует способа пополнить набор u_x, u_y и u_z четвертой компонентой u_t до какого-нибудь 4-вектора. В самом деле, если бы u_x, u_y, u_z, u_t были компонентами 4-вектора в S', то y-компонента в S' удовлетворяла бы условию

$$u_{y'} = u_y.$$

С другой стороны, $u_{y'}$ должна была бы быть y-компонентой обычной (классической) скорости, измеренной в системе S', и, следовательно,

$$u_{y'} = \frac{\sqrt{1-(v/c)^2}}{1 - uv/c^2}\, u_y,$$

вопреки тому, что $u_{y'} = u_y$. Это второе условие получается, если вспомнить, что $u_y = \Delta y/\Delta t$, где Δy — смещение в направлении оси y за время Δt. Подобным же образом

$$u_{y'} = \frac{\Delta y'}{\Delta t'} = \frac{\sqrt{1-(v/c)^2}\,\Delta y}{\Delta t - (v/c^2)\,\Delta x} = \frac{\sqrt{1-(v/c)^2}\, u_y}{1 - vu/c^2},$$

как и замечено выше.

С другой стороны, как легко проверить, четыре числа

$$\frac{u_x}{\sqrt{1-(u/c)^2}}, \quad \frac{u_y}{\sqrt{1-(u/c)^2}}, \quad \frac{u_z}{\sqrt{1-(u/c)^2}}, \quad \frac{c}{\sqrt{1-(u/c)^2}},$$

где $u^2 = u_x^2 + u_y^2 + u_z^2$ — квадрат классической скорости, уже являются компонентами некоторого 4-вектора. (Это выражение для 4-скорости подсказывается тем фактом, что для каждого 4-вектора w выражение $w_x^2+w_y^2+w_z^2-c^2w_t^2$ остается одним и тем же в любой лоренцевой системе отсчета.) Более того, для скоростей u, малых по сравнению со скоростью света c, пространственные компоненты нашего 4-вектора сводятся к компонентам u_x, u_y, u_z обычной скорости.

Теперь мы можем определить 4-импульс как 4-скорость, умноженную на собственную массу m; таким образом, его компонентами будут

$$\frac{mu_x}{\sqrt{1-(u/c)^2}},\quad \frac{mu_y}{\sqrt{1-(u/c)^2}},\quad \frac{mu_z}{\sqrt{1-(u/c)^2}},\quad \frac{mc}{\sqrt{1-(u/c)^2}}.$$

Итак, появился соблазн заменить классический закон сохранения импульса и энергии законом сохранения релятивистского 4-импульса. Этот релятивистский закон многократно подтверждался экспериментом — окончательным судьей любой физической теории; однако для нас здесь наиболее примечателен способ, посредством которого геометрические соображения открывают путь к глубокому описанию физической реальности.

Чтобы оценить всю глубину этого вновь открытого закона сохранения 4-импульса, посмотрим, что можно извлечь из сохранения четвертой компоненты.

Если отношение u/c мало, то (приближенно)

$$\frac{mc}{\sqrt{1-(u/c)^2}} \sim mc\left(1+\frac{1}{2}\frac{u^2}{c^2}\right)=\frac{1}{c}\left(mc^2+\frac{1}{2}mu^2\right).$$

Мы узнаем здесь величину $\frac{1}{2}mu^2$ — классическую кинетическую энергию. В классической трактовке явления столкновения сохранение энергии и сохранение импульса выступали как два совершенно разных закона. Геометрия связала воедино два эти закона сохранения! Но это еще не все. В формулу

$$\frac{mc}{\sqrt{1-(u/c)^2}} \sim \frac{1}{c}\left(mc^2+\frac{1}{2}mu^2\right),$$

кроме классической кинетической энергии $\frac{1}{2}mu^2$, входит еще совершенно новый энергетический член

mc^2. Это знаменитая эйнштейновская энергия покоя. И хотя для того, чтобы полностью понять ее смысл и оценить ее роль, требуется подробный физический анализ, невозможно не удивиться тому, что наличие энергии покоя в природе может быть обнаружено при помощи геометрии.

§ 17. Преобразования, потоки и эргодичность

Предположим, что E — евклидово пространство, а T — взаимно однозначное преобразование этого пространства на себя, т. е. отображение, переводящее E снова во все пространство E. Выберем некоторую точку p и рассмотрим точки

$$T^{-n}(p),\ T^{-n+1}(p),\ \ldots,\ p,\ T(p),\ T^2(p),\ \ldots,\ T^n(p),\ \ldots.$$

Как ведет себя эта (конечная или бесконечная) последовательность точек? Изучение таких последовательностей составляет ядро теории эргодических свойств преобразований. Ниже мы вкратце, на примерах, охарактеризуем возникающие здесь вопросы.

Допустим, что E — единичный куб или единичный шар в трехмерном пространстве, заполненный несжимаемой жидкостью, образующей стационарный поток в E. Поток называется стационарным, если в каждой точке пространства направление и скорость движения не зависят от времени. Если задан такой поток, можно рассмотреть преобразование $T(p)$, где p — любая точка нашего пространства, определенное так: $T(p)$ есть положение частицы жидкости, находящейся в данный момент в точке p, через одну секунду. Чтобы узнать положение исходной частицы через две секунды, нужно только найти точку $T^2(p)$, и, вообще, положение частицы через n секунд задается точкой $T^n(p)$. (Слова «одна секунда» означают здесь, конечно, произвольную единицу времени.) Предполагается, что поток несжимаем, т. е. если произвольная подобласть A в E имеет объем $m(A)$, то множество $T(A)$ точек, занимаемых по прошествии единицы времени частицами, находившимися первоначально в A, имеет тот же объем:

$$m(T(A)) = m(A) \quad \text{для всех} \quad A.$$

Такие потоки изучаются в гидродинамике, но они играют важную роль и в общей теории динамических систем. Там они появляются в связи со следующей интерпретацией задач механики. Если заданы динамическая система, состоящая из n материальных точек в трехмерном пространстве, начальные положения этих точек и их импульсы, то всю систему можно представить одной точкой $6n$-мерного пространства (потребуются три пространственные координаты и три компоненты импульса для каждой из n точек). Пространство этих точек называется фазовым пространством рассматриваемой системы. Величины и направления сил между точками (которые могут зависеть, например, только от взаимных расстояний между ними) задаются математически. С течением времени координаты и импульсы изменяются в соответствии с уравнениями динамики; при этом каждая «фазовая точка» движется в фазовом пространстве. Таким образом, мы получаем поток в $6n$-мерном пространстве. Как показал Лиувилль, для консервативных динамических систем этот поток несжимаем, т. е. сохраняет объем в фазовом пространстве.

Термин «консервативная» оэначает постоянство энергии, а это в свою очередь означает, что некоторая определенная функция координат и импульсов остается постоянной во время этого движения. Следовательно, на фазовую точку налагается ограничение, состоящее в том, что она должна двигаться по некоторой поверхности E постоянной энергии (называемой энергетической поверхностью), и эволюция динамической системы определяет поток на этой поверхности.

Сохранение обычного объема в полном фазовом пространстве *индуцирует* сохранение некоторой корректно (хотя и довольно сложно) определенной меры на энергетической поверхности. Следовательно, для большого семейства подмножеств A энергетической поверхности E существует счетно аддитивная мера m, такая, что

$$m(A) = m^*(T(A)).$$

Наблюдая за фазовой точкой через 1, 2, ..., n, ... секунд, мы придем к итерациям сохраняющего меру m на E преобразования T.

Важная гипотеза, сформулированная впервые Больцманом, состоит в том, что «в общем случае» (т. е. для большинства динамических систем) траектория фазовой точки с течением времени пройдет через все точки энергетической поверхности. В этом заключалась первоначальная «эргодическая гипотеза». Вскоре стало понятно, что это невозможно, причем невозможность была доказана на чисто топологической основе: «кривая» (т. е. взаимно однозначный образ бесконечной прямой) не может проходить через все точки энергетической поверхности размерности $k > 2$. Однако первоначально сформулированная гипотеза была слишком сильной: для целей больцмановской статистической механики хватило бы и более слабого свойства. В частности, достаточно было бы установить, что «в общем случае» кривая должна проходить произвольно близко к любой точке энергетической поверхности E. Иными словами, последовательность точек p, $T(p)$, $T^2(p)$, ..., $T^n(p)$, ... должна быть *плотной* в E. Хотелось бы узнать, так ли это для многих или даже для «большинства» *сохраняющих объем* преобразований T.

Существование преобразований T, сохраняющих объем и эргодических в указанном выше смысле, было установлено. Было даже показано, что преобразования, не обладающие этим свойством, в некотором корректно определенном смысле являются исключительными среди всех преобразований, сохраняющих меру. На самом деле можно доказать больше: последовательность итераций точки не только плотна в пространстве, но и *равномерно*-плотна. Это асимптотически равномерное поведение последовательности точек p, $T(p)$, ..., $T^n(p)$, ... можно определить так. Рассмотрим множество A меры $m(A)$ и попытаемся, начав с некоторой точки p, найти частоту, с которой ее итерации попадают в A. Эту частоту «попаданий» в A можно сжато выразить в символической записи. Определим характеристическую функцию $\chi_A(p)$ множества A, полагая $\chi_A(p) = 1$, если p принадлежит A, и $\chi_A(p) = 0$ в противном случае. Тогда искомая частота для итераций от 1 до N запишется в виде

$$\frac{1}{N}\sum_{i=1}^{N}\chi_A(T^i(p));$$

эргодическая теорема утверждает, что в пределе при $N \to \infty$ для почти всех точек p («почти всех» в смысле меры m) эта частота равна относительной мере области A:

$$\lim \frac{1}{N}\sum_{i=1}^{N}\chi_A(T^i(p)) = \frac{m(A)}{m(E)}.$$

Дж. Д. Биркгоф первым доказал, что рассматриваемый предел *существует* для почти всех p. Позднее было показано, что для многих преобразований этот предел действительно равен $m(A)/m(E)$. На самом деле можно показать, что в некотором определенном смысле *большинство* сохраняющих объем преобразований обладает этим свойством.

Итерационный процесс имеет смысл и тогда, когда отображение T (конечного множества целых чисел в себя, или непрерывное отображение отрезка в себя, или непрерывное отображение n-мерного пространства в себя) не является взаимно однозначным. В таком случае гораздо труднее определить свойства последовательности итераций отдельных точек. Имеются примеры, когда об их поведении еще что-то можно сказать. Возьмем, скажем, отображение интервала $(0, 1)$ на себя, определяемое формулой $x' = f(x) = 4x(1-x)$. Графиком этой функции служит парабола; рассматривая итерации функции $f(x)$, мы будем получать полиномы все более высокой степени, графики которых будут иметь возрастающее число максимумов и минимумов. В этом случае можно доказать, что последовательность итераций, начинающаяся с почти каждой точки, плотна на всем интервале (0, 1).

По счастливой случайности об этом преобразовании можно сказать гораздо больше. Полагая $x = \sin^2\theta$, мы получаем $f(x) = 4\sin^2\theta(1-\sin^2\theta) = \sin^2 2\theta$. Следовательно, отображение x в $4x(1-x)$ эквивалентно более простому отображению θ в 2θ.

Однако в общем случае даже для простых алгебраических отображений свойства итераций определить достаточно трудно.

Следует заметить, что эргодическую теорему можно распространить и на отображения, не являющиеся взаимно однозначными, если слова «сохраняющие меру» понимать в том смысле, что мера «прообраза» множества A (т. е. множества, которое преобразуется в A) должна быть равна мере A.

Эргодическая теорема для не взаимно однозначных отображений находит поразительное применение в теории непрерывных дробей. Пусть x — некоторое действительное число между 0 и 1. Чтобы представить x в виде непрерывной дроби, поступим так. Разделим 1 на число, обратное к x, и запишем это обратное к x число в виде суммы ближайшего к нему целого числа a_1 и неотрицательной дроби x_1. Затем проделаем все то же самое с числом x_1 и будем продолжать этот процесс до бесконечности или до тех пор, пока он не закончится (последнее произойдет в том случае, когда x — рациональное число).

Например,

$$\frac{17}{21} = \frac{1}{\frac{21}{17}} = \frac{1}{1+\frac{4}{17}} = \frac{1}{1+\frac{1}{\frac{17}{4}}} = \frac{1}{1+\frac{1}{4+\frac{1}{4}}}$$

или

$$\sqrt{2}-1 = \frac{1}{\sqrt{2}+1} = \frac{1}{2+(\sqrt{2}-1)} = \frac{1}{2+\frac{1}{\sqrt{2}+1}} =$$

$$= \frac{1}{2+\frac{1}{2+\frac{1}{2+\frac{1}{2+\dots}}}} = \dots.$$

Алгоритм разложения числа x в непрерывную дробь можно записать так: пусть

$$\frac{1}{x}=a_1+x_1, \quad \frac{1}{x_1}=a_2+x_2 \quad \text{и т. д.;}$$

тогда

$$x=\cfrac{1}{a_1+\cfrac{1}{a_2+\cfrac{1}{a_3+\dots}}}.$$

Определим теперь отображение T интервала $(0, 1)$ на себя формулой

$$Tx=\frac{1}{x}-\left[\frac{1}{x}\right],$$

$\left[\frac{1}{x}\right]$ — ближайшее к $\frac{1}{x}$ целое число, не превосходящее $\frac{1}{x}$.

Тогда

$$a_1(x)=\left[\frac{1}{x}\right], \quad a_2=a_1(Tx), \quad a_3=a_1(T^2x), \quad \dots,$$

и мы снова имеем дело с итерациями отображения T.

Прообразом интервала (a, b), где $0<a<b<1$, является объединение бесконечного множества интервалов

$$\left(\frac{1}{1+b}, \frac{1}{1+a}\right), \quad \left(\frac{1}{2+b}, \frac{1}{2+a}\right), \quad \left(\frac{1}{3+b}, \frac{1}{3+a}\right), \quad \dots$$

Положим, по определению, меру интервала (α, β) равной

$$\frac{1}{\log 2}\log\frac{1+\beta}{1+\alpha}=\frac{1}{\log 2}\int_{\alpha}^{\beta}\frac{dx}{1+x};$$

можно проверить, что тогда мера интервала (a, b) оказывается равной сумме мер интервалов, составляющих его прообраз. Отсюда следует, что если определить меру m подмножества A интервала $(0, 1)$ формулой

$$m(A)=\frac{1}{\log 2}\int_{A}\frac{dx}{1+x},$$

то преобразование T будет сохранять эту меру. Затем можно показать, что применима эргодическая теорема. При этом получается, например, следующий результат:

Для почти всякого x (в смысле меры m, или, что равносильно, в смысле обычной меры Лебега) частота, с которой целое число k появляется в последовательности $a_1, a_2, \dots$, равна

$$\frac{1}{\log 2}\log\left(\frac{(k+1)^2}{k(k+2)}\right).$$

Этот результат вытекает из эргодической теоремы в приведенной выше формулировке, если считать, что E — интервал $(0, 1)$, A — интервал $\left(1+\frac{1}{k+1},\ 1+\frac{1}{k}\right)$, т. е. множество всех x, для которых $a_1(x)=k$, а мера m — определенная выше логарифмическая мера.

Мы остановились столь подробно на этом примере для того, чтобы еще раз продемонстрировать чудо решительного вмешательства теории, возникшей в одном контексте, в иной, совершенно с ним не связанный. Эргодическая теория, появившаяся благодаря попыткам Больцмана поставить на солидную основу кинетическую теорию газов, нашла применение в теории непрерывных дробей.

§ 18. Еще об итерации и композиции отображений

Простейшими отображениями n-мерного евклидова пространства в себя являются линейные преобразования.

Как мы видели в § 15, такие преобразования имеют следующий общий вид:

$$\begin{aligned}
x_1' &= t_{11}x_1 + t_{12}x_2 + \ldots + t_{1n}x_n,\\
x_2' &= t_{21}x_1 + t_{22}x_2 + \ldots + t_{2n}x_n,\\
x_3' &= t_{31}x_1 + t_{32}x_2 + \ldots + t_{3n}x_n,\\
&\ldots\ldots\ldots\ldots\ldots\\
x_n' &= t_{n1}x_1 + t_{n2}x_2 + \ldots + t_{nn}x_n,
\end{aligned}$$

где t_{ij} — действительные числа. Как и раньше, обозначим матрицу коэффициентов через T. Особенно важен и интересен тот случай, когда все числа t_{ij} положительны или, точнее, неотрицательны[1]). Такие матрицы встречаются во многих приложениях в алгебре, теории вероятностей, теории распределения частиц в системах при цепной реакции, таких, как системы нейтронов в реакторах, и в математических моделях, применяемых в экономике. Если T — положительная матрица, то легко проверить, что векторы, заполняющие «положительный октант» n-мерного пространства, определяемый условиями $x_1 \geqslant 0$, $x_2 \geqslant 0$, ..., $x_n \geqslant 0$,

---

[1]) Положительность элементов матрицы T не является «внутренним» свойством линейного преобразования, ибо в другой системе координат его матрица, вообще говоря, теряет это свойство. Однако можно внутренним образом определить «положительное преобразование» как преобразование, переводящее в себя некоторое заданное неограниченное выпуклое множество (называемое «октантом»). Тогда можно показать, что существует система координат, в которой рассматриваемое преобразование представляется матрицей с неотрицательными элементами.

переводятся преобразованием T снова в векторы из того же октанта.

Рассмотрим теперь поверхность единичной сферы в нашем пространстве и, в частности, ее кусок, лежащий в положительном октанте, т. е. множество C всех векторов $(x_1, x_2, \ldots, x_n)$, компоненты которых неотрицательны и удовлетворяют уравнению $x_1^2 + x_2^2 + \ldots + x_n^2 = 1$. Определим другое преобразование S так: для вектора x, лежащего в положительном октанте, обозначим через x^* кратный ему единичный вектор и положим $S(x^*) = (T(x))^*$. Преобразование S переводит в себя часть единичной сферы, лежащую в положительном октанте, — множество, топологически эквивалентное $(n-1)$-мерному кубу. (В случае двумерного пространства это дуга окружности — топологически то же самое, что отрезок; в трехмерном случае это «восьмушка» сферы, топологически эквивалентная кругу, и т. д.) Мы уже установили раньше, что непрерывное отображение n-мерного куба или шара в себя должно иметь неподвижную точку (теорема Брауэра о неподвижной точке). Следовательно, на рассматриваемой части единичной сферы должна найтись такая точка $\overset{*}{x_0}$, что $S\left(\overset{*}{x_0}\right) = \overset{*}{x_0}$. Из определения преобразования S видно, что тогда в положительном октанте имеется вектор x_0, который переводится преобразованием T в вектор, кратный самому себе, т. е. такой, что $T(x_0) = \lambda x_0$. Так мы убеждаемся в существовании собственного вектора x_0 и соответствующего ему собственного значения λ линейного преобразования T. Можно также доказать, хотя это и труднее, что для матриц со строго положительными элементами такой вектор единствен и что если начать с любого вектора x из положительной области пространства и образовать последовательность $x, T(x), T^2(x), \ldots, T^n(x), \ldots$, то эти векторы будут сходиться *по направлению* к этому единственному собственному вектору. Эта теорема, доказанная впервые Фробениусом, имеет много интересных приложений.

Матрицы с неотрицательными элементами, пожалуй, наиболее широко используются в теории марковских цепей. (Марковские цепи — это обобщение понятия независимых испытаний.)

Часто бывает так, что система, которая может находиться в любом из состояний $s_1, s_2, s_3, \ldots$, в процессе своей эволюции переходит из одного состояния в другое, причем эти переходы происходят случайно. Можно считать, что переход занимает время τ. Переход из состояния s_i в состояние s_j имеет вероятность (основная, или одношаговая, вероятность перехода)

$$p_{ij} = \mathrm{Prob}\{s_i \to s_j \quad \text{за время} \quad \tau\};$$

при этом предполагается, что «сложная» вероятность, соответствующая последовательности n простых переходов (каждый из которых происходит за время τ)

$$s_i \to s_{i_1} \to s_{i_2} \to \ldots \to s_{i_{n-1}} \to s_j,$$

равна произведению соответствующих вероятностей:

$$p_{ii_1}p_{i_1i_2}\ldots p_{i_{n-1}j}.$$

Это «мультипликативное предположение» составляет основу понятия марковской цепи.

Чтобы привести пример марковской цепи, рассмотрим две урны I и II, в которых как-то распределены $2R$ занумерованных шаров. Каждые τ секунд мы случайно выбираем некоторое число от 1 до $2R$ (т. е. каждый раз выбор каждого из $2R$ чисел имеет вероятность $1/2R$) и перемещаем шар с этим номером из той урны, в которой он находился, в другую.

Можно считать, что состояние этой системы описывается числом i шаров в урне I. Тогда возможны лишь переходы $i \to i-1$ *или* $i \to i+1$, с тем исключением, что при $i=0$ возможен только переход $0 \to 1$, а при $i=2R$ — только переход $2R \to 2R-1$. Ясно, что одношаговые вероятности перехода равны

$$p_{ij}=0 \quad (\text{если } j \neq i-1,\ i+1),$$
$$p_{i,\,i-1}=i/2R,$$
$$p_{i,\,i+1}=1-i/2R;$$

«мультипликативное предположение» следует из независимости последовательных вытягиваний чисел от 1 до $2R$.

Эта модель была предложена Паулем и Татьяной Эренфестами в 1907 г. для того, чтобы проиллюстрировать некоторые логические трудности, возникающие при попытках примирить обратимые во времени законы динамики с необратимостью тепловых процессов, вытекающей из второго закона термодинамики.

К этой модели мы вернемся в гл. 3, а пока продолжим общее обсуждение понятия цепи Маркова.

Как вытекает из мультипликативного предположения и аксиомы аддитивности, вероятность того, что система, находившаяся первоначально в состоянии s_i, окажется после n переходов (т. е. через время $n\tau$) в состоянии s_j, равна (i, j)-му элементу n-й степени матрицы P вероятностей перехода. В самом деле, эта вероятность равна сумме по *всем* $i_1, i_2, \ldots, i_{n-1}$ величин вида

$$p_{ii_1}p_{i_1i_2}\ldots p_{i_{n-1}j},$$

а это и есть (i, j)-й элемент матрицы P^n. Итак,

$$\mathrm{Prob}\{s_i \to s_j \text{ за время } n\tau\}=(i,\ j)\text{-й элемент в } P^n,$$

где

$$P=\begin{pmatrix} p_{11} & p_{12} \ldots \\ p_{21} & p_{22} \ldots \\ \ldots & \ldots \end{pmatrix}.$$

Важно помнить, что умножение матриц (и, в частности, возведение матрицы в степень) было введено в непосредственной связи с композицией линейных преобразований и, в частности, итерациями одного линейного преобразования. Здесь же мы сталкиваемся с операцией возведения отображения в степень n в ситуации, не имеющей ничего общего с той, которая побудила нас ввести и изучить эту операцию. Но раз уж такое чудо случилось (а оно случается чаще, чем мы могли бы ожидать), можно в полной мере воспользоваться им и применить к изучению марковских цепей весь арсенал полученных в алгебре и геометрии сведений о матрицах.

С другой стороны, поскольку примеры марковских цепей часто возникают вне математики (например, в физике), мы получаем новые источники интуиции, которые могут подсказать, как обращаться с задачами, касающимися итерации матриц. Так, ведущая общая тема (в данном случае тема итерации) сближает и связывает ранее разобщенные области исследования к безмерной взаимной выгоде.

Итерация — это простейший пример композиции отображений, когда берутся только композиции отображения (или функции) с самим собой. Допустим, что заданы два преобразования S и T некоторого множества E в себя. Вместо группы, порожденной итерациями, рассмотрим всевозможные произведения S и T, а именно T^2, S^2, ST, S^2T, STS, TST и т. д., а также обратные им преобразования T^{-2}, S^{-2}, ... и все комбинации вида TS^{-1} или ST^{-1} или $TS^{-1}T$ и т. д. Короче говоря, рассмотрим полную группу, порожденную двумя преобразованиями S и T. Число элементов этой группы по-прежнему счетно, однако в нее могут входить элементы весьма разнообразного вида.

Можно доказать, что, исходя из двух взаимно однозначных непрерывных преобразований T и S отрезка в себя и рассматривая всевозможные их композиции, можно получить всюду плотный класс непрерывных преобразований отрезка в себя. Иными словами, если заданы произвольное непрерывное преобразование R и число $\varepsilon > 0$, то найдется произведение P, составленное из конечного числа преобразований T, S, T^{-1} и S^{-1}, такое, что $|P(x) - R(x)| < \varepsilon$ для всех x, т. е. любое непрерывное преобразование можно сколь угодно точно аппроксимировать преобразованиями нашего класса.

Аналогичная теорема справедлива и для пространств высших размерностей: если E есть n-мерная сфера, то можно найти конечное число (в действительности достаточно четырех) гомеоморфизмов E на себя (гомеоморфизм — это взаимно однозначное непрерывное отображение на себя), таких, что при помощи их композиций можно сколь угодно точно аппроксимировать любой заданный гомеоморфизм. Доказательство этой теоремы таково, что трудно сказать, какими свойствами обладают эти гомеоморфизмы. Было бы полезно показать, например, что эти гомеоморфизмы можно выбрать из класса всюду дифференцируемых отображений (если бы кому-нибудь удалось это сделать). В таком случае была бы решена одна из остающихся до сих пор открытыми проблем топологии: всякий ли гомеоморфизм n-мерного пространства можно аппроксимировать дифференцируемыми гомеоморфизмами.

Следует подчеркнуть, что эта задача аппроксимации касается гомеоморфизмов, т. е. взаимно однозначных непрерывных отображений. Если не требовать взаимной однозначности, то ответ будет положительным, ибо всякая непрерывная функция в замкнутой области n-мерного пространства по теореме Вейерштрасса может быть аппроксимирована многочленом, и, следовательно, всякое непрерывное отображение может быть аппроксимировано дифференцируемым.

Еще несколько задач, касающихся композиций отображений, покажут, как быстро можно достичь границ неизвестного. Пусть E — евклидова плоскость. Рассмотрим группу всех гомеоморфизмов, получающихся путем композиции гомеоморфизмов вида

$$x' = f(x, y), \quad y' = y \qquad \text{и} \qquad x' = x, \quad y' = g(x, y).$$

Оказывается, любой гомеоморфизм $x' = \varphi(x, y)$, $y' = \psi(x, y)$ можно аппроксимировать элементами нашей группы. В случае размерности 3 и более аналогичные задачи еще не решены. Например, в трехмерном пространстве можно взять порождающие отображения в виде $x' = f(x, y, z)$, $y' = y$, $z' = z$ и два аналогичных или рассмотреть группу G, порожденную всеми гомеоморфизмами вида $x' = f(y, z)$, $y' = g(x, z)$, $z' = h(x, y)$. В обоих случаях вопрос о том, аппроксимируем ли произвольный гомеоморфизм n-мерного пространства гомеоморфизмами рассматриваемого типа, все еще остается открытым.

В заключение упомянем недавние прекрасные результаты А. Н. Колмогорова и В. И. Арнольда по представлению произвольных действительных непрерывных функций любого числа переменных в виде суперпозиций подобных же функций, но уже только двух переменных. Оказывается, непрерывные функции многих переменных можно не приближенно, а *точно* выразить в виде суперпозиций лишь конечного числа функций двух переменных, например:

$$f(x_1, x_2, x_3, x_4) = h_1(h_2(x_1, x_2), h_3(x_3, x_4))$$

или

$$f = h_1(h_2(h_3(x_1, x_2), h_4(x_3, x_4)))$$

и т. д.

§ 19. Легко ли доказать очевидное?

В этом параграфе мы поговорим подробнее об отношениях между элементарной интуицией и математической строгостью. Мы возьмем примеры главным образом из топологии и попытаемся на них показать, как геометрическая интуиция позволяет в определенных случаях обнаруживать математические факты, которые можно доказать лишь при помощи тщательно разработанных и иногда очень сложных методов. С другой стороны, существуют интуитивно правдоподобные утверждения, которые *не верны*, хотя найти опровергающие их примеры довольно трудно.

Одна из важных теорем топологии евклидовой плоскости утверждает, что простая замкнутая кривая

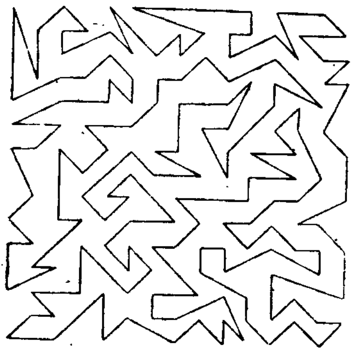


Рис. 24. Какие точки плоскости лежат внутри области, ограниченной этой кривой, а какие вне нее?

делит плоскость на две области — внутреннюю и внешнюю по отношению к этой кривой. Простая замкнутая кривая — это взаимно однозначный непрерывный образ окружности, т. е. на окружности определено непрерывное отображение, такое, что двум различным точкам окружности соответствуют две различные точки плоскости. Тогда множество точек образа *разделяет* плоскость, т. е. на плоскости существуют два непересекающихся множества точек, которые нельзя соединить дугой так, чтобы она не пересекала этот образ (кривую). Это утверждение кажется совершенно очевидным, однако строго доказать его не так уж просто; впервые это удалось сделать Камиллу Жордану. В самом деле, если кривая очень сильно изогнута (рис. 24) или делает много витков подобно туго накрученной спирали, то на глаз довольно трудно отличить внутреннюю область от внешней. (На сельских ярмарках раньше предлагалась такая игра: закручивали сложенные вдвое бечевку или кожаный ремень и соединяли вместе концы; зритель должен был, воткнув карандаш между двумя смежными кусками, угадать, где окажется карандаш — внутри или снаружи, когда ремень будет распрямлен. Ясно, что нечестный демонстратор всегда мог выиграть, соединив концы тем или иным способом.)

Объект типа «простой замкнутой кривой» олицетворяет понятие, не настолько ясное в своей полной общности, насколько прозрачны примеры, породившие его, ибо эти примеры в большинстве случаев представляют собой выпуклые или лишь умеренно изогнутые линии. Мы вернемся к этому в главе 2.

Вот еще одна теорема, которая может показаться очевидной и которую, однако, совсем нелегко доказать. Попытаемся представить себе непрерывное распределение векторов на поверхности сферы, т. е. приклеить в каждой точке сферы коротенький отрезок фиксированной длины, касательный к сфере и направленный так, чтобы при движении по сфере направления измеиялись непрерывно. Оказывается, это невозможно: сферу иельзя «причесать»! На ней всегда найдется некоторая точка, в которой вектор уже не будет касательным, — так сказать, «макушка». Эта невозможность следует из топологической теоремы Брауэра: для всякого непрерывного преобразования сферы в себя либо существует пара неподвижных точек, либо некоторая точка переходит в диаметрально противоположную. Невозможность «причесать» сферу, казалось бы, совершенно очевидна интуитивио, однако не известно никакого простого и в то же время полного доказательства этого факта.

Рассмотрим другую теорему о сфере. Допустим, что каждой точке сферы поставлены в соответствие два действительных числа, причем это соответствие непрерывно; иначе говоря, на поверхности сферы определены две действительные непрерывные фуикции $f_1(p)$ и $f_2(p)$. Тогда существует по крайней мере одна точка p_0, такая, что в ней и в диаметрально противоположной ей точке p_0^* обе функции принимают одни и те же значения:

$$f_1(p_0) = f_1(p_0^*),$$
$$f_2(p_0) = f_2(p_0^*).$$

В качестве таких функций можно рассматривать, скажем, температуру и давление в каждой точке поверхности Земли (считая ее сферической). Если предположить, что эти функции непрерывны, то на Земле должна найтись такая точка, где температура и давление такие же, как в диаметрально противоположной точке. Эта теорема имеет несколько забавных следствий; одно из них известно под названием «теоремы о бутерброде с ветчиной». Если заданы любые три тела в пространстве, то существует плоскость, разбивающая каждое из них на две равновеликие части. (Этими телами могут быть ломоть хлеба, слой масла и кусок ветчины, — отсюда и идет название теоремы.) Наметим вкратце доказательство этой теоремы, опирающееся на теорему о диаметрально противоположных точках (которую доказать не так просто). Рассмотрим ориентированные плоскости в пространстве, т. е. плоскости, для каждой из которых определено направление «положительной нормали» к плоскости. В таком случае множество «направлений» плоскостей, указываемых выбором «положительной нормали», задается как множество лучей, проходящих через центр некоторой фиксированной сферы. Эту сферу мы назовем сферой направлений. В любом выбранном направлении можно провести плоскость, разделяющую первое из трех тел на равновеликие части. (Это можно вывести из того, что части, на которые тело делится плоскостью фиксированного направления, непрерывным образом зависят от положения плоскости.) Посмотрим теперь, как разобьет эта плоскость два других тела. Обозначим через $f_1(p)$ разность между объемами «положительной» и «отрицательной» частей, на которые она разобьет второе тело; здесь «положительной» считается часть тела, расположенная с той стороны от плоскости, которую указывает «положительная» нормаль. (Разумеется, тело может находиться целиком по одну сторону от плоскости, и тогда разность будет равняться полному объему тела, взятому с надлежащим знаком.) Аналогично определяется функция $f_2(p)$ для третьего тела. Эти две функции определены и непрерывны в каждой точке P сферы направлений. Следовательно, существует точка, в которой они принимают те же значения, что и в диаметрально противоположной точке. Но при переходе к диаметрально противоположным точкам обе функции, как следует из их определения, меняют знак, ибо при этом меняется ориентация делящей плоскости. Поскольку единственное число, равное своему отрицанию, — это 0, отсюда следует наше утверждение.

ГЛАВА 2

ТЕМЫ, ТЕНДЕНЦИИ И СИНТЕЗ

Вероятно, наиболее поразительная черта математики как интеллектуальной дисциплины — это громадное разнообразие проблем, которыми она занимается. Например, задача о числе способов размена доллара и задача построения $\sqrt[3]{2}$ циркулем и линейкой кажутся очень далекими, и невозможно не удивляться той глубокой смысловой связи, которая существует между ними.

Это разнообразие проблем сочетается с отсутствием четких критериев, очерчивающих предмет математики, поэтому здесь чрезвычайно трудно добиться сколько-нибудь полного синтеза и унификации. Приходится, кроме того, остерегаться унификации столь общей, что она уже стала бы тривиальной, и избегать слишком жесткого синтеза, который ограничивал бы будущий рост и развитие науки. Следует упомянуть, что единственная за последние годы серьезная попытка представить всю математику в целом с единой точки зрения, предпринятая группой Бурбаки, подвергалась критике с обеих этих позиций[1]).

До середины 19-го века было мало сознательных попыток, направленных на достижение синтеза или унификации. Конечно, существовали «Элементы» Евклида — наиболее полный синтез, и декартова аналитическая геометрия — наиболее разработанная уни-

---

[1]) По поводу общих установок группы французских математиков, пишущих под коллективным псевдонимом Н. Бурбаки, см., например, книгу Н. Бурбаки «Очерки по истории математики», М., ИЛ, 1963, стр. 245—249. О группе Бурбаки см. статью П. Халмоша «Николай Бурбаки» (сборник «Математическое просвещение», вып. 5, М., Физматгиз, 1960, стр. 229—239). — *Прим. ред.*

фикация алгебры и геометрии; однако математики после Ньютона были слишком заняты, с радостью устремившись к новым далям, которые открыли им алгоритмы дифференциального и интегрального исчисления, и у них просто не хватало времени наводить порядок в своих быстро расширяющихся владениях.

Затем наступила реакция и появилась тенденция к организации добытых знаний, которая продолжается до сих пор. Одной из причин этой перемены было то, что «тело» математики очень сильно выросло и потребовалась какая-то система жестких связей, без которой отдельные его части оказались бы полностью разобщенными. Кроме того, математиков начинала беспокоить не связанная никакими ограничениями интуиция, которой не препятствовали стандарты, налагаемые формальной системой.

Когда-то модель строгости, созданная Евклидом, была непревзойденной. Однако по мере того, как математикам открывался постоянно расширяющийся поток проблем, их критическое чутье обострилось, а логическое стало более тонким и точным.

Для математика 18-го века тот факт, что простая замкнутая кривая разделяет плоскость на две части, был настолько очевиден, что едва ли заслуживал упоминания. Но в 19-м веке Жордан, уловивший, в чем здесь тонкость, уже попытался, хотя и не вполне успешно, доказать этот факт. (По-настоящему простого доказательства теоремы Жордана о кривых нет даже и сейчас.)

Математик 18-го или начала 19-го века пользовался понятием простой замкнутой кривой чисто интуитивно. Ко временам Жордана под простой замкнутой кривой стали понимать результат *непрерывного взаимно однозначного отображения окружности*. Поскольку окружность, очевидно, разделяет плоскость на внутреннюю и внешнюю области, ясно, казалось бы, что и непрерывный взаимно однозначный образ окружности обязательно сохранит это свойство. Вероятно, это представляется нам столь очевидным потому, что мы наделяем непрерывность всеми видами свойств. Поскольку наша интуиция не только «питается», но и лимитируется физической реальностью, слова «простая замкнутая кривая» вызывают в воображении сравнительно гладкую кривую, возможно, с несколькими острыми выступами, — в худшем случае нечто вроде кривой, изображенной на рис. 25.

Однако «непрерывный взаимно однозначный образ окружности» (*формальное* определение простой замкнутой кривой) может оказаться совершенно «диким»

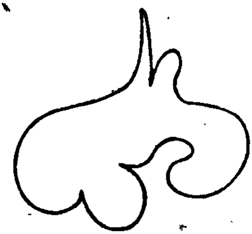


Рис. 25.

множеством. Например, это может быть кривая бесконечной длины или, еще хуже, кривая, не имеющая касательной *ни в одной* из своих точек!

Первый пример такой кривой дал Вейерштрасс; она описывается параметрическими уравнениями

$$x = \sin\theta,$$

$$y = \sum_{n=1}^{\infty} \frac{1}{2^n} \cos 3^n \theta.$$

Ясно, что x — непрерывная функция от θ. Воспользовавшись тем, что коэффициенты $\frac{1}{2^n}$ бесконечного ряда, определяющего y, убывают со скоростью геометрической прогрессии, а $\cos 3^n\theta$ всегда заключен между -1 и $+1$, можно показать, что y — тоже непрерывная функция от θ. Таким образом, если θ меняется от 0 до 2π, то соответствующая точка $(x(\theta), y(\theta))$ непрерывно движется вдоль некоторой кривой. Для того чтобы эта кривая имела касательную в некоторой точке, соответствующей какому-то значению θ, должны существовать обе производные $dx/d\theta$ и $dy/d\theta$.

Вейерштрасс доказал, что $dy/d\theta$ не существует ни при каком θ. Доказать это нелегко, однако нетрудно заподозрить, ибо формальное (почленное) дифференцирование ряда для y приводит к ряду

$$\sum_{n=1}^{\infty}\left(\frac{3}{2}\right)^n \sin 3^n\theta,$$

который *расходится* при всех (кроме счетного множества) значениях θ.

Отбросив требование взаимной однозначности, можно построить (первым это сделал Пеано) непрерывный образ отрезка (т. е. кривую, правда, не являющуюся простой), заполняющий квадрат. Подумав о таких патологических созданиях, получающих право на существование, если постулируется одна только непрерывность, начинаешь понимать, почему Жордану понадобилось доказывать свою теорему и почему ее доказательство совсем не простое.

Для «гладких» кривых, к которым касательную можно провести «почти» в каждой точке (т. е. в каждой за исключением самое большее конечного числа точек), теорема Жордана доказывается намного легче.

Лишь около 1830 г. понятия непрерывности, гладкости и им подобные получили более или менее ясные определения; до этого ими пользовались неформально, а иногда и бессистемно.

Напомним, что середина 19-го века явилась началом новой эры в математике, которая характеризуется растущим недоверием к интуиции, не подкрепленной доказательствами, и усилившимися надеждами на логику. В результате математика стремится выглядеть более строгой, более формальной, черпающей критерии и методы внутри самой себя. Ничто не принимается на веру, и ничто не может избежать тщательной проверки. Даже Евклид был подвергнут исчерпывающему логическому анализу, и в созданном им великолепном сооружении обнаружились трещины. Например, Евклид пренебрег целой группой аксиом, необходимых для формализации понятия «лежать между», — так называемыми аксиомами порядка. Эти аксиомы казались столь очевидно и столь тривиально «истинными», что Евклид и его последователи считали их само собой разумеющимися. Однако полная формализация означает, что геометрии можно учить слепого и даже вычислительную машину. Многие рассуждения Евклида опираются на использование того, что некоторая точка D на прямой, проведенной через точки A и B, лежит между этими точками. Стандартное доказательство равенства углов CAB и CBA в равнобедренном треугольнике ABC $(AC = BC)$ построено на том, что из C опускается перпендикуляр на AB; он пересекает AB в точке D, и для завершения доказательства требуется только установить, что D лежит *между* A и B. Это невозможно «объяснить» ни слепому, ни вычислительной машине, не формализовав понятие «лежать между». Более полную аксиоматику геометрии дал в 1899 г. Гильберт в своих знаменитых «Основаниях геометрии»[1]).

Понятие числа также подверглось тщательному анализу, что способствовало развитию новой алгебры и самой логики.

В результате математическая алгебра стала тем, чем она главным образом является сейчас: учением о таких абстрактных системах, как группы, кольца, поля. Рассмотрим следующий пример, показывающий, что такое рассуждение в стиле современной алгебры и как этот стиль проникает в те ветви математики, которые традиционно считались далекими от алгебры.

Начнем с предположения, что мы знаем, что такое целые числа (положительные, отрицательные и нуль).

Целые числа можно складывать, вычитать (т. е. в их области разрешимы уравнения вида $a + x = b$) и умножать. Эти операции над целыми числами обладают следующими свойствами:

[1]) Даже аксиоматика Гильберта может доставить требовательному логику достаточно поводов для беспокойства, в первую очередь из-за аксиомы непрерывности, включенной Гильбертом в систему аксиом. Эта аксиома, устанавливающая взаимно однозначное и сохраняющее порядок соответствие между точками прямой и действительными числами, сразу же переносит в геометрию все трудности, связанные с несчетностью множества действительных чисел.

1) $a+b=b+a$;
2) $a+(b+c)=(a+b)+c$;
3) $ab=ba$;
4) $(ab)c=a(bc)$;
5) существует одно и только одно целое число, а именно 0, такое, что $a+0=a$ для любого целого a;
6) существует одно и только одно целое число, а именно 1, такое, что $a\cdot 1=a$ для любого целого a;
7) не существует делителей нуля, т. е. из $ab=0$ следует, что либо $a=0$, либо $b=0$ (либо $a=b=0$).

Далее для построения класса рациональных чисел поступают следующим образом.

Рассмотрим все упорядоченные пары целых чисел, в которых второе число отлично от нуля, и назовем две пары $(a;\ b)$ и $(c;\ d)$ *эквивалентными*: $(a, b) \sim (c, d)$, если $ad=bc$.

Определенное так отношение эквивалентности обладает следующими свойствами:

(а) $(a, b) \sim (a, b)$ (рефлексивность);

(b) из $(a, b) \sim (c, d)$ следует $(c, d) \sim (a, b)$ (симметричность);

(c) из $(a, b) \sim (c, d)$ и $(c, d) \sim (e, f)$ следует $(a, b) \sim (e, f)$ (транзитивность).

Свойства (а) и (b) кажутся настолько простыми, что не требуют пояснений; (c) следует из свойства 7) выше. В самом деле, из $ad=bc$ и $cf=ed$ получаем (умножая первое равенство на e, а второе — на a)

$$ade=bce \quad \text{и} \quad cfa=eda,$$

откуда $bce=cfa$, или $(be-fa)c=0$. Следовательно, либо $c=0$ и тогда $a=0$ и $e=0$, поскольку $b\neq 0$, $d\neq 0$, либо $be=af$.

Рассмотрим теперь весьма общий принцип, широко применяемый в математике.

Пусть S — некоторое множество объектов α, β, γ, ..., связанных некоторым бинарным отношением R, рефлексивным ($\alpha R\alpha$ для всех α), симметричным (из $\alpha R\beta$ следует $\beta R\alpha$) и транзитивным (из $\alpha R\beta$ и $\beta R\gamma$ следует $\alpha R\gamma$). Тогда S разбивается на попарно не пересекающиеся классы, такие, что объекты из одного и того же класса связаны отношением R, а объекты из разных классов — не связаны.

Пусть, например, α, β, γ, ... — множества, а $\alpha R \beta$ означает, что между элементами α и элементами β имеется взаимно однозначное соответствие. Тем самым между множествами устанавливается бинарное отношение, которое, очевидно, рефлексивно, симметрично и транзитивно. Совокупность всех множеств распадается на непересекающиеся классы, состоящие из множеств одинаковой мощности.

Уайтхед и Рассел определяют кардинальные числа как классы множеств одинаковой мощности. Так, число «три» — это класс всех множеств, состоящих из трех элементов, $\aleph_0$ (алеф нуль) — класс всех счетных множеств и т. д. Хотя отождествление чисел с соответствующими классами эквивалентности — это уже довольно высокая степень абстракции, некоторые ратуют за то, чтобы именно так вводить числа в младших классах школы.

Возвращаясь к нашим упорядоченным парам, мы видим, что, поскольку отношение эквивалентности $\sim$ обладает тремя требуемыми свойствами, множество всех пар также распадается на непересекающиеся «классы эквивалентности».

Отождествим упорядоченную пару (a, b) $(b \neq 0)$ с дробью a/b; тогда введенное отношение эквивалентности выражает не что иное, как условие равенства дробей. При этом множество всех упорядоченных пар разбивается на классы эквивалентности, в каждый из которых собраны все дроби, *представляющие* одно и то же рациональное число.

Но можно ли так говорить, когда мы еще не определили, что понимается под рациональным числом? Нельзя; однако поскольку на самом деле мы знаем, о чем идет речь, мы можем сделать все эти рассуждения совершенно законными, *отождествив* рациональные числа с соответствующими классами эквивалентности. Иными словами, *рациональные числа суть по определению классы эквивалентных между собой пар целых чисел.*

Пусть r и s — рациональные числа, т. е. r (а также и s) есть класс эквивалентных пар.

Чтобы определить их сумму $r+s$, возьмем некоторую пару (a, b) из r (так называемый представитель класса r) и некоторую пару (c, d) из s и построим пару $(ad+bc, bd)$. (Заметим, что $ad+bc$ — это числитель, а bd — знаменатель дроби, полученной сложением a/b и c/d по обычным правилам.) Тогда $r+s$ — это класс пар, эквивалентных паре $(ad+bc, bd)$. Может показаться, что это определение зависит от выбора представителей в r и s. Однако можно проверить, что если (a', b') — другая пара из r, т. е. $(a, b) \sim (a', b')$, или $a'b = ab'$, и (c', d') — пара из s, то

$$(a'd' + b'c', \; b'd') \sim (ad + bc, \; bd)$$

и, следовательно, пара $(a'd' + b'c', b'd')$ принадлежит классу $r+s$. Аналогично можно определить rs как класс, содержащий пару (ac, bd), где (a, b) — пара из r, а (c, d) — из s.

А как обстоит дело с исходными целыми числами? Они теперь выступают в несколько замаскированной форме как классы, содержащие пары вида $(a, 1)$. Иначе говоря, целое число a — это класс пар, эквивалентных паре $(a, 1)$.

Уравнение $ax = b$ не всегда разрешимо в целых числах (например, уравнение $2x = 3$ не имеет целых решений); в области же рациональных чисел при $a \neq 0$ оно всегда имеет решение. В самом деле, x — это просто класс пар, эквивалентных паре (b, a). Более того, любое уравнение $rx = s$, где $r \neq 0$ и r, s — рациональные числа, разрешимо в рациональных числах.

То, что мы сделали, можно охарактеризовать как *вложение* множества целых чисел в большее множество (а именно в множество рациональных чисел), причем таким способом, что в этом большем множестве стала возможной операция деления (деление на нуль, конечно, исключается). При таком вложении мы сохранили целые числа и операции над ними и в то же время разумным образом распространили эти операции на все рациональные числа.

Итак, мы описали общую формальную схему расширения системы объектов, в которой можно выполнять операции сложения и умножения, в систему, в которой стало возможным еще и деление. При этом объектами необязательно должны быть целые числа, а сложение и умножение — это необязательно обычные арифметические операции. Все, что требуется, — это выполнение приведенных выше условий 1) — 7), *каковы бы ни были объекты и операции.* На самом деле некоторые из этих условий излишни; так, можно обойтись без условия 6). С другой стороны, условие 7) весьма существенно.

Проиллюстрируем теперь преимущества нашего формального подхода на другом примере.

В качестве объектов будем рассматривать *непрерывные функции* $a(t)$, $b(t)$, ... действительной переменной t, определенные при $0 \leqslant t < \infty$. Сложение определяется обычным образом, а в качестве умножения берется так называемая *свертка*:

$$a * b = \int_0^t a(t-\tau)\, b(\tau)\, d\tau = \int_0^t b(t-\tau)\, a(\tau)\, d\tau = b * a.$$

Например, если $a(t) = 1$ и $b(t) = t$, то $a * b = t^2/2$; если $a(t) = t$ и $b(t) = \sin t$, то $a * b = t - \sin t$, и т. д. Стоит отметить, что свертка с $a(t) \equiv 1$ эквивалентна интегрированию от 0 до t. Таким способом *трансцендентная* операция интегрирования приобретает «облик» *алгебраической* операции умножения. И, как мы покажем в нескольких последующих абзацах, наблюдаемая здесь аналогия отнюдь не является только «внешней». Операция свертки часто встречается в разных вопросах чистой и прикладной математики и потому широко изучалась. Давно было замечено, что во многих отношениях свертка напоминает обычное умножение.

В самом деле, математик легко проверит свойства 1) — 5). Свойство 6) не выполняется (но, как упоминалось выше, это ни на что не влияет); свойство 7) можно доказать, но это уже отнюдь не просто. Свойство 7) обычно характеризуют словами: в множестве непрерывных функций относительно операции свертки не существует делителей нуля.

Далее можно поступить в точности так же, как выше, и расширить рассматриваемое множество функций до множества классов эквивалентных пар функций, сделав таким способом уравнение $a * x = b$ всегда разрешимым в пределах этого большего множества.

При этом аналогами рациональных чисел оказываются определенные операторы, по существу совпадающие с теми, которые ввел Оливер Хевисайд для решения линейных дифференциальных уравнений, встречающихся в теории электрических цепей.

Например, уравнение

$$\int_0^t x(\tau)\,d\tau = 1, \qquad t \geqslant 0,$$

не имеет решения, являющегося непрерывной функцией от t. Однако формально мы можем написать

$$x = \frac{\{1\}}{\{1\}},$$

где $\{1\}$ — функция, принимающая значение 1 при всех $t \geqslant 0$. Под делением понимается, конечно, не обычное деление, а операция, обратная свертке. Тогда

$$x * f = \frac{\{1\}}{\{1\}} * f = \frac{\{1 * f\}}{\{1\}} = \frac{\int_0^t f(\tau)\,d\tau}{\{1\}} = f,$$

так что x — это просто *оператор умножения на число* 1.

Взяв более общее уравнение

$$1 * x = \int_0^t x(\tau)\,d\tau = a(t),$$

можно показать, что x есть оператор s, равный сумме оператора дифференцирования D и оператора $a(0)\dfrac{\{1\}}{\{1\}}$, где $a(0)$ — значение функции a при $t = 0$:

$$sa = Da + a(0)\frac{\{1\}}{\{1\}}.$$

Обозначая через p оператор интегрирования

$$pb = \int_0^t b(\tau)\,d\tau = 1 * b,$$

можно проверить, что

$$sp = ps = \text{тождественный оператор},$$

т. е.

$$(sp)\,a = (ps)\,a = a$$

для любой дифференцируемой функции a.

При таком способе определения работа с операторами становится совершенно аналогичной манипуляциям с обычными дробями; следует только помнить, что здесь умножение — это операция свертки.

Этот способ введения операторов по аналогии с введением рациональных чисел придуман сравнительно недавно польским математиком Яном Микусинским. Использование же операторов для формального решения линейных дифференциальных уравнений было известно уже давно; начало этому положил, как мы уже упоминали, Хевисайд в конце 19-го века[1].

Подход Микусинского к операторному исчислению Хевисайда — отличный пример того, что можно назвать алгебраизацией математики. Это продукт тенденции, начавшейся в 19-м веке и продолжающейся

---

[1] Некоторые из современников Хевисайда критиковали его за использование формальных приемов без ясного понимания их содержания и смысла. Говорят, что в ответ своим критикам Хевисайд как-то сказал: «Должен ли я отказаться от хорошего обеда лишь потому, что не понимаю процессов пищеварения?» Подобным образом можно было бы критиковать шестиклассника, который учится пользоваться дробями, не понимая лежащей в основе теории.

Мы остановились на этом потому, что здесь хорошо видна одна из сильных тенденций современной математики: игнорировать и отвергать всё, что не формализовано логически. Именно эта тенденция (начинающая проникать в начальное и среднее обучение) в большой степени ответственна за растущее отделение математики от физики. Физик, применяющий матемав наши дни с небывалой силой и напором,— подогнать математику под шаблоны, заготовленные абстрактными алгебраическими структурами.

Наиболее заметный успех алгебраизация имела в топологии. Примером может служить созданная Эмилем Артином теория кос.

Пусть l_1 и l_2 — две параллельные и одинаково направленные прямые в пространстве (см. рис. 26; направления прямых указаны стрелками).

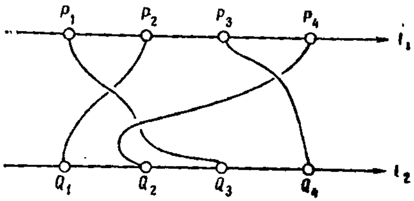


Рис. 26.

Выберем n различных точек $P_1, P_2, \ldots, P_n$ на l_1, занумерованных в соответствии с направлением l_1; аналогично выберем n точек $Q_1, Q_2, \ldots, Q_n$ на l_2. Для простоты и определенности мы будем всюду считать $n = 4$.

Каждая точка P_i соединена с некоторой точкой Q_j кривой c; эта кривая может изгибаться и закручиваться в пространстве, однако мы требуем, чтобы ее проекция на плоскость, определяемую прямыми l_1 и l_2, была монотонной, т. е. чтобы при движении точки R вдоль этой проекции от P_i до Q_j ее расстояние от прямой l_1 монотонно росло.

Не допускается, чтобы какие-нибудь две кривые пересекались в пространстве; в частности, две кривые

тические методы, вполне может положиться на внутреннюю согласованность своих построений и, что самое важное, на совпадение полученных результатов с экспериментом. Подобно шестикласснику, он будет рад воспользоваться рациональными числами, не зная во всех подробностях, как их можно объединить в формальную систему, и, подобно Хевисайду, будет счастлив жонглировать операторами, не дожидаясь, когда логика даст ему разрешение на это.

не могут кончаться в одной и той же точке Q_j. На рисунке, чтобы показать, какая из кривых «выше» другой, мы прерываем в соответствующем месте проекцию той из них, которая «ниже».

Все, что мы делали до сих пор, было описанием объекта, который можно было бы назвать *сплетением*.

Чтобы определить *косу*, мы должны еще ввести класс *деформаций*, которые, хотя и способны самым решительным образом изменить *внешний вид* сплетения, сохраняют, однако, его существенные черты.

Эти деформации обладают следующими свойствами:

(a) прямые l_1 и l_2 остаются параллельными и одинаково направленными, хотя расстояние между ними может как угодно увеличиться или уменьшиться;

(b) точки P и Q могут как угодно передвигаться вдоль соответствующих прямых, однако их порядок сохраняется;

(c) никакие две кривые не могут пересечься в процессе деформации: они «неразрывны»;

(d) кривые могут как угодно растягиваться или сжиматься, однако для их проекций на плоскость (l_1, l_2) сохраняется указанное выше свойство монотонности.

Эти свойства станут совершенно понятными, если представить себе, что l_1 и l_2 — жесткие стержни, а кривые c — гибкие резиновые шнуры, которые можно как угодно растягивать.

Назовем два сплетения *эквивалентными*, если одно из них можно превратить в другое при помощи деформации, удовлетворяющей перечисленным выше четырем условиям. Это отношение эквивалентности обладает основными свойствами рефлексивности, симметричности и транзитивности; следовательно, все сплетения распадаются на попарно не пересекающиеся классы эквивалентности.

Коса — это и есть такой класс эквивалентности.

Главная проблема теории кос — разработать процедуру (алгоритм), которая позволит решать, совпадают или нет две косы (или, что то же самое, эквивалентны или нет два сплетения).

Эта задача относится к геометрии, точнее к ветви геометрии, известной под названием топологии. Топологию можно определить как учение о тех свойствах геометрических конфигураций, которые не меняются (инвариантны) при определенного вида *непрерывных* деформациях. Решается же эта задача чисто алгебраическими средствами: путем определения некоторой специальной группы и отождествления каждой косы с элементом этой группы.

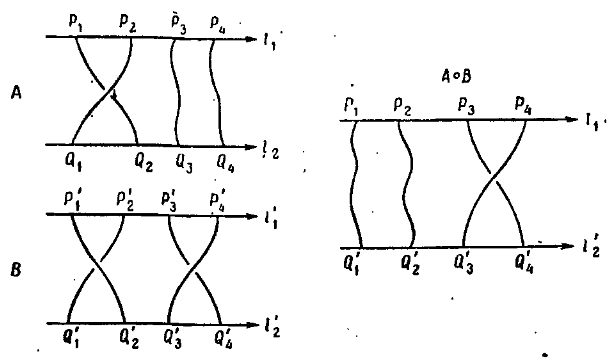


Рис. 27.

Чтобы показать, как это делается, определим сначала операцию *композиции* кос.

Пусть A и B — две косы (для обеих $n=4$); чтобы построить косу $A \circ B$, выберем по одному сплетению из A и из B (рис. 27).

Прямые, точки и кривые для A обозначим через $(l_1, l_2, P_1, P_2, P_3, P_4, Q_1, Q_2, Q_3, Q_4, c_1, c_2, c_3, c_4)$, а для B — через $(l'_1, l'_2, P'_1, P'_2, P'_3, P'_4, Q'_1, Q'_2, Q'_3, Q'_4, c'_1, c'_2, c'_3, c'_4)$.

Продеформируем B так, чтобы плоскость (l'_1, l'_2) совпала с плоскостью (l_1, l_2), а прямая l'_1 — с прямой l_2 (включая ориентацию); позаботимся еще о том, чтобы l_1 и l'_2 лежали по разные стороны от l_2.

Продолжим деформацию и добьемся того, чтобы точка P'_1 совпала с Q_1, P'_2 — с Q_2 и т. д. После этого отбросим (сотрем) l_1', а значит, и l_2, «связав» подходящим образом кривые c_i и c_i'. Полученное сплетение и есть представитель класса, определяющего косу $A \circ B$.

Простой пример композиции кос показан на рис. 27. В этом примере $A \circ B = B \circ A$; однако, вообще говоря, это равенство не выполняется.

Тем не менее операция композиции кос *ассоциативна*:

$$(A \circ B) \circ C = A \circ (B \circ C)$$

и существует единственный единичный элемент I — тривиальная коса, представленная на рис. 28.

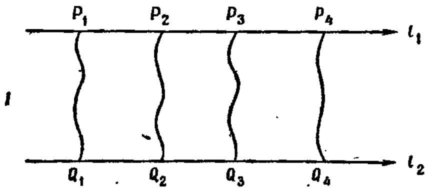


Рис. 28.

Нетрудно заметить, что для любой косы A

$$A \circ I = I \circ A = A.$$

Всякая коса A имеет также единственную *обратную* A^{-1}, т. е. такую косу A^{-1}, что

$$A^{-1} \circ A = A \circ A^{-1} = I.$$

Чтобы по данной косе A построить косу A^{-1}, достаточно поменять местами свойства «выше» и «ниже», а именно, если в A кривая c_i проходит выше c_j, то в A^{-1} она должна проходить ниже, и наоборот; все остальное не меняется.

Таким образом, мы видим, что относительно операции композиции косы образуют группу.

В этой группе имеются три элемента (три косы), которые вместе со своими обратными играют особенно важную роль. Они представлены на рис. 29. (Можно проверить, что, например, $A_1 \circ A_2 = A_2 \circ A_1$.)

Эта важная роль состоит в том, что любая коса может быть представлена в виде композиции этих «базисных» кос.

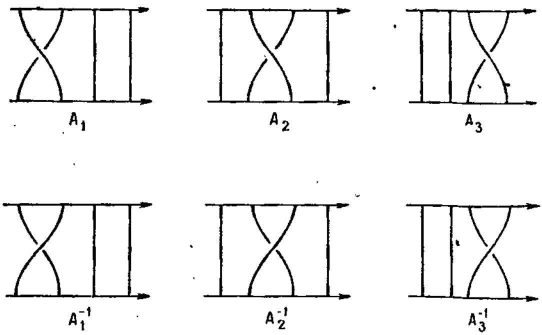


Рис. 29.

Например, сплетение, приведенное на рис. 26, можно деформировать так, что оно будет выглядеть, как на рис. 30. Пунктирная линия на этом рисунке

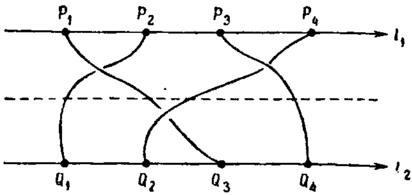


Рис. 30.

проведена для того, чтобы легче было увидеть, что соответствующую косу можно записать как

$$A_1 \circ A_2^{-1} \circ A_3.$$

Обозначим элементы A_1^{-1}, A_2^{-1}, A_3^{-1} соответственно через B_1, B_2, B_3. В этих обозначениях каждое сплетение выражается «словом» типа

$$A_2 \circ B_1 \circ B_2 \circ B_3 \circ A_3 \circ A_1.$$

Два сплетения эквивалентны тогда (и только тогда), когда соответствующие «слова» представляют *один и* *тот же* элемент группы. Например, $A_1 \circ B_1$ и $A_2 \circ B_2$ — разные слова, но они представляют один и тот же элемент группы, а именно, единичный элемент I.

Кроме очевидных соотношений

(a) $$A_1 \circ B_1 = A_2 \circ B_2 = A_3 \circ B_3 = I,$$

между порождающими буквами (или образующими группы) имеется еще только одно соотношение

(b) $$A_1 \circ A_3 = A_3 \circ A_1.$$

Другим независимым соотношением могло бы показаться, например, соотношение $B_1 \circ B_3 = B_3 \circ B_1$, однако его можно вывести из (a) и (b).

Теперь можно забыть о косах и сформулировать задачу в чисто алгебраических терминах: задана некоторая группа, порождаемая шестью образующими A_1, A_2, A_3, B_1, B_2, B_3, удовлетворяющими соотношениям (a) и (b); существует ли алгоритм, устанавливающий представляют ли два слова один и тот же элемент группы?

Для кос при любом n такой алгоритм известен и сравнительно несложен. Однако тесно связанная с этой проблема классификации узлов, хотя и приводит к родственной рассмотренной выше задаче эквивалентности слов для аналогично построенной группы, до сих пор остается нерешенной.

Теория кос иллюстрирует два важных обстоятельства.

Во-первых, она показывает, что задача, относящаяся по своей природе к области *непрерывного*, может быть успешно решена методами дискретного и комбинаторного характера. (Другой подобный пример — использование комбинаторной леммы Шпернера в доказательстве теоремы Брауэра о неподвижной точке; см. § 6 гл. 1.)

Во-вторых, задача может потребовать изобретения новой символики или нового формального аппарата. Если бы теория групп была неизвестна, могло бы случиться (хотя это в высшей степени маловероятно), что ее придумали бы специально для решения проблемы классификации кос или другой подобной задачи.

В этом и состоит математическое творчество — либо понять, что к данной задаче применим уже существующий аппарат, либо изобрести новый.

При всем сказанном выше, если бы потребовалось назвать одного-единственного человека, работы которого наиболее решительным образом повлияли на современный дух математики, то большинство голосов почти наверняка собрал бы Георг Кантор. К середине 19-го века накопление материала достигло в математике таких размеров и такого разнообразия, что настало время обратиться к синтезу и пересмотру оснований. Так родилась математическая теория множеств, с одной стороны, и теория математических систем (математическая логика, метаматематика) — с другой.

В частности, математики стали с большей ответственностью относиться к вопросам *строгости* при введении понятий и построении доказательств. Проводимое с этой точки зрения исследование оснований анализа (исчисления бесконечно малых) и попытки понять смысл системы действительных чисел и установить общие свойства функций от них привели к задачам, положившим начало современной теории множеств и современной математической логике.

Великие аналитики и геометры после Ньютона (например, Бернулли, Эйлер, Даламбер, Лагранж) обладали почти безошибочным инстинктом, формулируя верные теоремы и давая правильные доказательства без твердой опоры на формальные системы и без точного соблюдения стандартов логической строгости. Вряд ли можно сомневаться в том, что математическая *интуиция* (служа гению) обеспечивает такую ясность и такое единство, что она предвосхищает любой специальный формализм и делает его практически излишним.

Какова природа и каков источник математической интуиции — это проблемы философии и психологии. Вероятно, в далеком будущем, когда станет более понятным устройство нервной системы и мозга человека, некоторый свет прольется и на такие вопросы. Если окажется, что характер логического мышления в математике в большой степени определяется этим устройством и находится под его влиянием, то, возможно, станет яснее, почему интуитивная математика допускает столь естественную формализацию, причем по существу единственным способом.

Как создатель любой великой новой теории, Кантор имел предшественников. Уже Галилей заметил, что натуральных чисел столько же, сколько квадратов натуральных чисел, потому что между этими двумя множествами можно установить взаимно однозначное соответствие. Джордано Бруно явно считал физические объекты бесконечными. Однако только в середине 19-го века математики Больцано, Вейерштрасс, Дедекинд и логики Буль, де Морган и несколько позднее Фреге и Пеано поставили вопросы и построили системы, вошедшие в современное здание теории множеств. Но своей полной общности и размаха эта теория достигла лишь благодаря Кантору.

Кантор ввел множества неформально и не использовал для установления их свойств никакой аксиоматики. Он положил в основу понятие взаимно однозначного соответствия, а в качестве примеров рассматривал множества, с которыми в то время сталкивались математики в своих исследованиях. К общим идеям его привело, по-видимому, изучение рядов Фурье. Он заметил, что множество всех целых чисел, множество всех рациональных чисел и множество всех алгебраических чисел имеют одинаковую мощность, т. е. между ними можно установить взаимно однозначные соответствия. Затем был установлен важнейший результат: мощность континуума действительных чисел больше мощности счетного множества. Он был доказан с помощью знаменитого диагонального метода Кантора — простого, но сыгравшего весьма важную роль во всей теории. Затем последовали работы, посвященные строению точечных множеств и новому созданию Кантора — порядковым числам. В этих работах содержались начала топологии точечных множеств, теории функций действительной переменной, построение трансфинитных порядковых чисел, критическое обсуждение оснований дифференциального исчисления (с отказом от бесконечно малых), критическое и философское (скорее чем аксиоматическое) обсуждение природы континуума и оснований математики вообще.

К концу своего творчества Кантор пришел к другим важным выводам, кульминацией которых явилась постановка проблемы континуума. Грубо говоря, эту знаменитую проблему Кантора можно сформулировать так: существует ли множество, более мощное, чем множество всех целых чисел, но менее мощное, чем множество всех действительных чисел?

Теория множеств развивалась в действительности по двум различным направлениям, хотя, к счастью, в мышлении Кантора они были тесно объединены.

Одно направление разрабатывало понятие мощности и другие понятия чисто *абстрактным* образом, т. е. безотносительно к природе рассматриваемых множеств. Второе касалось точечных множеств на прямой, на плоскости или в многомерных евклидовых пространствах.

Первый подход вскоре слился с логикой; второй же породил топологию точечных множеств и привел в конце концов к весьма плодотворной теории абстрактных пространств, положив начало весьма важной тенденции к геометризации математики.

Покажем на примере, что мы называем геометризацией математики. Рассмотрим сначала числа вида

$$\frac{\varepsilon_1}{2}+\frac{\varepsilon_2}{2^2}+\frac{\varepsilon_3}{2^3}+\ldots,$$

где каждое ε_i — либо 0, либо 1. Чтобы избежать повторения некоторых чисел, например

$$\frac{1}{2}=\frac{1}{2}+\frac{0}{2^2}+\frac{0}{2^3}+\ldots=\frac{0}{2}+\frac{1}{2^2}+\frac{1}{2^3}+\ldots,$$

условимся использовать в подобных случаях то из двух представлений, которое оканчивается бесконечной последовательностью нулей (т. е. первое из выписанных выше представлений числа $^1/_2$). Затем рассмотрим множество чисел вида

$$\frac{2\varepsilon_1}{3}+\frac{2\varepsilon_2}{3^2}+\frac{2\varepsilon_3}{3^3}+\ldots.$$

Ясно, что эти два множества находятся во взаимно однозначном соответствии и, следовательно, имеют одну и ту же мощность; они даже имеют один и тот же порядковый тип: естественное взаимно однозначное соответствие между ними сохраняет порядок точек в обоих множествах (т. е. отношение «больше или равно»). Таким образом, абстрактно эти множества неразличимы, хотя как точечные множества на прямой они кажутся совершенно различными.

Первое множество — это весь отрезок (0, 1), включая его концы, а второе — знаменитое канторовское множество, которое получается из отрезка (0, 1) выбрасыванием его открытой средней трети (т. е. всех точек между $^1/_3$ и $^2/_3$, кроме самих этих точек) и затем последовательным (до бесконечности) выбрасыванием открытых средних третей оставшихся отрезков.

Канторовское множество, хотя и имеет мощность континуума, очень разреженное. Оно *нигде не плотно* на отрезке (0, 1), т. е. каждую не принадлежащую ему точку можно окружить интервалом, тоже не содержащим точек канторовского множества. Понятию *нигде не плотного* множества противопоставляется понятие множества, *всюду плотного* в интервале. Множество называют *всюду плотным* в некотором интервале, если *каждый* подинтервал содержит хотя бы одну точку (а следовательно, бесконечно много точек!) этого множества. Рациональные числа между 0 и 1 образуют всюду плотное множество в интервале (0, 1), и все же это множество, будучи счетным, имеет меньшую мощность, чем канторовское множество.

Итак, мы видим, что одно множество может быть мощным, но разреженным, а другое — скудным, но плотным. Ясно, что здесь перед нами два совершенно разных понятия. Кардинальное число, или мощность, — это обобщение понятия *количества элементов*; оно связано со счетом, в то время как разреженность и плотность связаны с *расположением в пространстве* и понятием «близости».

Объединяя эти два понятия, можно получить вполне удовлетворительное определение «малых» множеств. Множество называют *множеством первой категории* («малым»), если оно является объединением (не более чем) *счетного числа нигде не плотных множеств*. При этом и множество рациональных чисел, и канторовское множество оказываются «малыми»: первое — потому что оно скудно, второе — потому что разреженно.

Но теперь можно пойти гораздо дальше. Проанализировав понятие плотности, легко обнаружить, что на самом деле здесь «работает» понятие *окрестности точки*, обобщающее понятие интервала. В свою очередь в основу этого понятия можно положить понятие расстояния. (Интервал с центром x_0 можно определить как множество точек x, расстояние которых до x_0 не превосходит некоторого заданного положительного числа δ; в многомерных пространствах таким способом в качестве аналогов интервала вводятся шары.)

Оказывается, что для многих целей существенны лишь три свойства расстояния:

(a) расстояние между a и b неотрицательно; оно равно нулю тогда и только тогда, когда a и b совпадают;

(b) расстояние от a до b равно расстоянию от b до a;

(c) расстояние между a и b не превосходит суммы расстояний от a до c и от b до c, каково бы ни было c (так называемое неравенство треугольника).

Рассмотрим, например, множество C всех непрерывных функций действительной переменной t, определенных при $0 \leqslant t \leqslant 1$. Если заданы две такие функции $a(t)$ и $b(t)$, то «расстояние» между ними $\delta(a, b)$ можно определить формулой

$$\delta(a, b) = \max_{0 \leqslant t \leqslant 1} |a(t) - b(t)|;$$

легко проверить, что при этом выполняются основные условия (a), (b), (c).

Теперь многие чисто аналитические факты о непрерывных функциях можно выразить в геометрических терминах. Например, знаменитую теорему Вейерштрасса о том, что всякую непрерывную функцию можно с произвольной точностью приблизить многочленами, можно сформулировать по-новому, сказав, что множество многочленов всюду плотно в пространстве C непрерывных функций.

Такие на вид чисто словесные преобразования одних предложений в другие оказались чрезвычайно полезными как источники новых (геометрических) открытий и новых проблем.

Примером может служить замечательный результат Банаха о том, что непрерывные функции, имеющие производную хотя бы в одной точке, образуют в пространстве C всех непрерывных функций множество первой категории. Иными словами, непрерывные нигде не дифференцируемые функции — это вовсе не какие-нибудь патологические создания; этим свойством обладает *подавляющее большинство* непрерывных функций. (В самом деле, такие функции образуют множество второй категории, т. е. дополнение в C «разреженного» множества первой категории.)

Вероятно, еще более замечательно то, что легче доказать, что «большинство» непрерывных функций нигде не дифференцируемы, чем привести пример хотя бы одной такой функции[1]).

Таким образом, понятие множества первой категории стало мощным средством доказательства существования определенных математических объектов.

Развитию аксиоматического метода способствовали не только фундаментальные работы Кантора, но и некоторые другие мотивы. Открытие неевклидовых геометрий и дальнейшее их развитие дали толчок к аксиоматической трактовке геометрических систем, отличных от евклидовой геометрии, включая и такие глубоко разработанные системы, как проективная

[1]) Аналогично, легче доказать, что «большинство» действительных чисел трансцендентны, чем построить хотя бы одно трансцендентное число (ср. § 4 гл. 1). — *Прим. ред.*

геометрия. Упомянутая выше работа Гильберта и работы американских геометров, например Веблена, придали новый интерес и размах использованию аксиоматического метода в других разделах математики. Работы Пеано по аксиоматизации *арифметики* и Буля по аксиоматизации алгебры множеств развивались по существу параллельно исследованиям по аксиоматическим основаниям геометрии[1]). После создания теории множеств можно и даже нужно было подумать о построении системы аксиом для всей математики. Такие попытки поощрялись элегантностью аксиоматического метода и его успешным применением в отдельных областях математики. И действительно, исследования, подобные работам Уайтхеда и Рассела[2]), во многом обязаны своим появлением опыту, накопленному при работе с аксиоматическими системами в некоторых разделах геометрии, арифметики и алгебры.

Гильберт выдвинул свою великую программу — заложить аксиоматический фундамент, на котором можно было бы базировать любые математические исследования. Это не означает, что существовало единодушие по поводу допустимости или смысла всех аксиом! В частности, некоторые математики считали аксиому выбора подозрительной и, по-видимому, неприемлемой из-за ее странных и, казалось бы, парадоксальных следствий. (Мы уже упоминали об одном таком парадоксе — разбиении сфер разных радиусов на одно и то же конечное число попарно конгруэнтных подмножеств.) Дебаты по поводу роли и смысла этой аксиомы продолжались и в начале 20-го столетия. Следует сказать, что на протяжении всей истории математики постоянно происходило открытие новых объектов, обладающих свойствами, которые в тот период были непривычными для математического мышления. Это

---

[1]) Эти работы предшествовали во времени исследованиям Д. Гильберта (и тем более О. Веблена). — *Прим. ред.*

[2]) Имеются в виду фундаментальные «Принципы математики» (Principia mathematica, 3 тома, 1910—1913) А. Н. Уайтхеда и Б. Рассела — одно из первых сочинений по основаниям математики. — *Прим. ред.*

случалось даже и без применения аксиом, устанавливающих (подобно аксиоме выбора) существование «неконструктивных» объектов. Процесс обобщения в математике очень часто начинался именно с таких «удивительных» открытий. Их логические следствия, какими бы странными они ни казались в момент зарождения, в конце концов принимались и часто оказывались основой новых систем. Школа интуиционистов, первоначально возглавляемая Л. Брауэром, а затем некоторое время А. Лебегом и Г. Вейлем, пыталась ограничить математику более конструктивными, или «операциональными», системами. Однако громадное большинство математиков не отвергает аксиому выбора.

Программа Гильберта означала веру в полноту некоей всеобъемлющей аксиоматической системы для всей математики. Работы Бернайса, Френкеля и фон Неймана уже заложили прочный фундамент аксиоматических оснований теории множеств и математической логики. Можно было надеяться, что все имеющие смысл проблемы в таких системах (в принципе) разрешимы.

Затем в 1931 г. Гёдель опубликовал свою статью «Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme I» (О формально неразрешимых предложениях Principia Mathematica и родственных систем I). Грубо говоря, его результат состоял в том, что в любой достаточно богатой системе аксиом (на самом деле — достаточно богатой, чтобы включать в себя арифметику) можно сформулировать предложения, которые имеют смысл и тем не менее неразрешимы внутри данной системы. Более того, некоторые из этих предложений оказываются верными, или «истинными», в следующем смысле: они имеют вид утверждений о том, что все целые числа обладают определенными арифметическими свойствами, выполнение которых можно проверить для каждого конкретного числа.

То, что «истинное» предложение бывает недоказуемым, можно проиллюстрировать таким примером. Рассмотрим предложение: *для всякого* натурального

n имеет место формула

$$1+2+\ldots+n=\frac{n(n+1)}{2}.$$

Обычно это доказывают, применяя аксиому математической индукции, которая гласит: если некое утверждение $S(n)$ о натуральных числах справедливо для $n=1$ и если из справедливости $S(m)$ следует справедливость $S(m+1)$, то $S(n)$ верно для *всех* n. Эта аксиома позволяет нам устанавливать некоторые предложения, касающиеся бесконечного множества объектов (в данном случае — натуральных чисел), не проверяя их бесконечно много раз, что было бы, разумеется, невыполнимой задачей. Если бы в нашем распоряжении не было аксиомы индукции, предложение, что *для каждого* n

$$1+2+\ldots+n=\frac{n(n+1)}{2},$$

нельзя было бы доказать, ибо среди всех аксиом арифметики и логики *одна лишь* аксиома индукции позволяет делать утверждения о всей бесконечной совокупности натуральных чисел.

Можно было бы попытаться сбросить ярмо индукции и рассуждать так: если бы для некоторого n сумма $1+2+\ldots+n$ не равнялась числу $n(n+1)/2$, то существовало бы наименьшее такое n; оно не могло бы равняться 1, поскольку наше утверждение для $n=1$ верно; но оно не могло бы быть и больше 1, ибо тогда можно было бы показать, что $n-1$ тоже исключительное число, а это противоречит тому, что n — наименьшее из таких чисел. Увы, это рассуждение основано на принципе, утверждающем, что каждое непустое множество натуральных чисел имеет наименьший элемент, а этот принцип *равносилен* аксиоме индукции.

Итак, без аксиомы индукции простые арифметические утверждения вроде

$$1+2+\ldots+n=\frac{n(n+1)}{2},$$

даже несмотря на то, что они «истинны», нельзя было бы вывести из остальных аксиом; можно сказать, что без аксиомы индукции арифметика была бы *неполной*.

Гёдель показал, что на самом деле *всякая* достаточно богатая система аксиом является неполной и не может быть сделана полной добавлением никакого конечного числа новых аксиом.

Подробно остановиться здесь на доказательстве Гёделя невозможно, однако мы приведем его краткий обзор. Гёдель занумеровал все допустимые математические предложения, в которых используются предписанные символы и правила операций данной системы; такие предложения образуют счетный класс, и Гёдель приписал каждому из них определенное целое число. Таким способом каждое математическое предложение было сведено к некоторому утверждению о целых числах. При помощи процесса, аналогичного диагональному методу Кантора (упомянутому выше), Гёдель в рамках рассматриваемой системы сформулировал предложение, которое нельзя ни доказать, ни опровергнуть средствами самой этой системы. (Эта теорема применима, разумеется, только к *непротиворечивым* системам, т. е. таким, в которых нельзя доказать противоречивое утверждение типа $1 = 0$.)

Чтобы пояснить чуть полнее, какие идеи лежат в основе конструкции Гёделя, мы должны сначала обсудить понятие *формальной системы*. Формальная система состоит из конечного множества символов и конечного числа правил, по которым эти символы можно объединять в формулы или предложения. Некоторые из этих предложений рассматриваются как аксиомы; повторное применение правил системы позволяет получать все новые и новые доказуемые предложения. Доказательство некоторого данного предложения (формулы) представляет собой конечную последовательность (цепочку) предложений, начинающуюся с одной из аксиом и заканчивающуюся требуемым предложением. Каждое промежуточное предложение из этой цепочки либо является аксиомой, либо выводится при помощи правил данной системы из предшествующих предложений.

Однако предложение, утверждающее, что некоторая последовательность формул образует (или не образует) доказательство некоей формулы, уже не является предложением в самой этой формальной системе. Это предложение *о системе*; такие предложения обычно называют *метаматематическими*.

Необходимо тщательно различать *математические* и *метаматематические* предложения. Нарушение этого условия приводит к парадоксам. Самый ранний из них принадлежит Эпимениду Критскому, высказавшему утверждение «все критяне лжецы», которое, очевидно, не может быть ни истинным, ни ложным.

Для наших целей важнее следующий вариант знаменитого парадокса Жюля Ришара.

Назовем функцию f, определенную для всех неотрицательных целых $n = 0, 1, 2, \ldots$ и принимающую неотрицательные целые значения, *вычислимой*, если для любого n существует некое предписание, состоящее из конечного числа слов и позволяющее посредством конечного числа шагов вычислить $f(n)$ (т. е. значение f при заданном n); это предписание может зависеть (и на самом деле почти всегда зависит) от n. Легко видеть, что множество всех вычислимых функций счетно. Следовательно, все вычислимые функции можно расположить в виде последовательности $f_1, f_2, f_3, \ldots$.

Определим теперь новую функцию g формулой

$$g(n) = f_n(n) + 1.$$

Она не входит в нашу последовательность, поскольку при $n = 1$ она отличается от f_1, при $n = 2$ — от f_2 и т. д. Следовательно, она *не вычислима*.

С другой стороны, ясно, что она вычислима, так как $f_n(n)$ вычислима, а прибавив 1 к $f_n(n)$, мы получаем $g(n)$.

Источник этого парадокса достаточно ясен: построение функции g существенно опирается на упорядочение функций f. Сами эти функции описываются в рамках некоторой системы (например, арифметики), однако их упорядочение — это уже *метаматематическая операция*.

Теперь представим себе, что метаматематическое предложение «m есть значение, которое принимает n-я функция последовательности f_1, f_2, ... в точке n» удалось каким-нибудь образом перевести в чисто арифметическое предложенне, т. е. предложение, законное в пределах рассматриваемой системы. Тогда, если эта система непротиворечива, предложение о вычислимости g будет неразрешимым.

У Гёделя возникла великая идея перевести метаматематические предложения в предложения арифметики, так сказать, *отразить* их внутрь формальной системы. При этом появилась возможность свободно комбинировать внутри системы математические и метаматематические предложения, и вопросы, которые при обычном ходе событий приводнли к парадоксам, превращались при таком отражении в *неразрешимые предложения*.

Несколько слов о гёделевской нумерации.

В формальной системе, включающей в себя арифметику, формула состоит из некоторого фиксированного набора знаков, таких, как $\Rightarrow$ — знак импликации, 0 — нуль, $\vee$ — логическое «или», (— левая скобка, а также *числовых переменных* x, y, z, ... (вместо них можно подставлять неотрицательные целые числа), *пропозициональных переменных* p, q, r, ..., заменяющих высказывания (предложения), и *предикатных переменных* P, Q, R, ... (вместо них можно подставлять такие предикаты, как «композиция» илн «больше чем»).

Можно обойтись десятью фиксированными знаками; им приписываются номера от 1 до 10 (скажем, знак $\Rightarrow$ получает номер 3, знак 0 — номер 6 и т. д.).

Числовым переменным сопоставляются простые числа, большие 10 (скажем, x получает номер 11, y — номер 13 и т. д.); пропозициональным переменным приписываются квадраты простых чисел, больших 10 (например, p имеет номер 11^2, q — номер 13^2 и т. д.); наконец, предикатным переменным соответствуют кубы простых чнсел, больших 10.

Формула получает номер по правилу, которое лучше всего пояснить на примере. Рассмотрим формулу

$$(x > y) \Rightarrow (x = sy) \vee (x > sy),$$

которая утверждает следующее: если x больше, чем y, то либо x следует непосредственно за y (т. е. $x = y + 1$), либо x больше, чем число, следующее непосредственно за y. Эта формула содержит 19 символов: (, x, $>$, y,), $\Rightarrow$, (, x, $=$, s, y,), $\vee$, (, x, $>$, s, y,), причем некоторые из них (скажем, x или s)

повторяются. Возьмем первые 19 простых чисел, возведем каждое из них в степень, равную номеру соответствующего символа, и перемножим полученные числа. Найденное так число

$$2^8 \times 3^{11} \times 5^{13^3} \times 7^{13} \times 11^9 \times 13^3 \times 17^8 \times 19^{13} \times 23^5 \times 29^7 \times$$
$$\times 31^{13} \times 37^9 \times 41^2 \times 43^3 \times 47^{11} \times 53^{13^3} \times 57^7 \times 61^{13} \times 67^9$$

и будет гёделевским номером этой формулы. Читателю следует заметить, что левой скобке (приписано число 8, правой скобке) — число 9, знаку $\vee$ — число 2, знаку s — число 7 и предикату $>$ — число 13^3.

Последовательность формул (которая может составлять доказательство) F_1, F_2, F_3, ..., F_m получает гёделевский номер

$$2^{G_1} \times 3^{G_2} \times \ldots \times p_m^{G_m},$$

где p_m есть m-е простое число, а G_1, G_2, ... — гёделевские номера соответственно формул F_1, F_2, ...

Таким способом каждой формуле или последовательности формул ставится в соответствие единственное натуральное число. Не всякое натуральное число является гёделевским номером, но всякий гёделевский номер *однозначно* определяет соответствующее выражение. Это следует из теоремы о единственности разложения на простые множители: для целого числа, большего 1, существует единственный (с точностью до порядка) способ записать его в виде произведения степеней простых чисел.

Заметим, что метаматематическое предложение «$(x > y)$ есть начальная часть формулы $(x > y) \Rightarrow (x = sy) \vee (x > sy)$» отражается внутрь системы, переходя в чисто арифметическое предложение: гёделевский номер формулы $(x > y)$, равный

$$2^8 \times 3^{11} \times 5^{13^3} \times 7^{13} \times 11^9,$$

является *делителем* гёделевского номера полной формулы.

Гёделевские конструкции, вообще говоря, приводят к предложениям *внутри системы*, допускающим толкование как предложения *о системе*. Выше мы говорили, что такие предложения могут, если не принять мер предосторожности, привести к парадоксам. Гёдель показал, как принятие таких мер превращает эти парадоксы в *неразрешимые предложения*.

Понятно, что открытие Гёделя произвело революцию в математической логике. Однако оно имело и гораздо более серьезные последствия, ибо повлекло за собой глубокое изменение философского взгляда на математику в целом.

Чтобы оценить это, следует понять, что гёделевские неразрешимые предложения не были какими-то тайными, известными лишь посвященным, утверждениями, стоящими где-то далеко в стороне от главного русла математики. Напротив, идея нумерации позволила формулировать их в терминах *диофантовых уравнений*, которые веками были вполне «благонамеренными» объектами чисто математических исследований.

Диофантовыми называют обычные алгебраические уравнения с одним или более неизвестными, для которых ищутся решения в целых числах.

С диофантовыми уравнениями связаны многие известные нерешенные задачи. Вероятно, самая знаменитая из них — теорема Ферма: при $n > 2$ уравнение

$$x^n + y^n = z^n$$

не имеет решения в целых числах. При $n = 2$ существует бесконечно много таких решений — это так называемые пифагоровы тройки (3, 4, 5), (5, 12, 13) и т. д. Куммер доказал теорему Ферма для $3 \leqslant n \leqslant 100$; получено также много родственных результатов.

Возможно ли, что эта проблема неразрешима в существующей системе математики?

Такой вопрос не мог бы и возникнуть до того, как Гёдель сделал свое основополагающее открытие. Благодаря Гёделю логика была поднята со своего привычного места в основаниях математики и включена во многие повседневные ее проблемы.

Существуют, вообще говоря, два вида математических доказательств: экзистенциальные и конструктивные. В главе 1 мы рассмотрели примеры обоих типов. Так, можно воспользоваться канторовским методом и доказать существование трансцендентных чисел, не приведя ни одного примера, а можно применить конструкцию Лиувилля и построить целый класс *конкретных* трансцендентных чисел.

Подобным же образом можно с помощью метода доказательства *от противного* установить, что всякое алгебраическое уравнение степени n с комплексными коэффициентами имеет хотя бы один комплексный корень; однако от такого доказательства мало проку, если интересоваться численным значением этого корня.

Эта возможность доказывать существование объектов без необходимости *предъявлять* их — одна из самых заметных отличительных черт математики. Однако чисто экзистенциальные рассуждения при свободном их использовании позволяют договориться до таких вещей, которые смутят кого угодно.

Мы уже упоминали, что некоторые совсем невинные аксиомы, вроде аксиомы выбора (если задана совокупность неперекрывающихся непустых множеств, то можно образовать новое множество, выбирая по одному элементу из каждого множества данной совокупности), позволяют доказывать существование очень странных объектов (например, неизмеримых множеств), резко противоречащих нашей интуиции.

Пожалуй, единственное методологическое разногласие среди математиков касается их отношения к *существованию* математических объектов. Однако все без исключения согласны, что *алгоритмы*, или *конструктивные процедуры*, весьма важны и эффективны.

Наиболее общий алгоритм можно определить в терминах так называемых рекурсивных функций. Вместо этого мы опишем один нетривиальный частный алгоритм в надежде, что это позволит выявить существенные черты всех алгоритмов.

Нашим примером будет алгоритм Евклида для нахождения целочисленных решений диофантова уравнения $ax + by = 1$, где a и b — неотрицательные целые числа.

Сначала заметим, что если a и b имеют общий делитель $d > 1$, то уравнение не имеет решений. (В противном случае левая часть делилась бы на d; тогда 1 должна была бы делиться на $d > 1$, что невозможно.) Итак, будем считать, что a и b взаимно просты, т. е. не имеют общих делителей, кроме 1.

Далее заметим, что если a или b равно 1, то мы немедленно получаем решение: $x = 1$, $y = 0$ (при $a = 1$) или $x = 0$, $y = 1$ (при $b = 1$).

В случае когда a и b взаимно просты и ни одно из них не равно 1 (откуда, в частности, следует, что $a \neq$

$\neq b$), алгоритм Евклида состоит из следующего набора указаний:

(1) Предположить, что a больше, чем b, и разделить a на b; получится частное q_1 и остаток r_1, *меньший*, чем b. Иначе говоря,

$$a = q_1 b + r_1, \quad 1 \leqslant r_1 < b.$$

Подставить полученное выражение в исходное уравнение:

$$r_1 x + b(q_1 x + y) = 1.$$

Полагая $x = x_1$ и $q_1 x + y = y_1$, находим, что

$$r_1 x_1 + b y_1 = 1.$$

(2) Заметить, что новое уравнение имеет тот же вид, что и старое, но при этом r_1 меньше, чем a (ибо r_1 меньше, чем b, а b меньше, чем a).

Разделить b на r_1, т. е. написать

$$b = q_2 r_1 + r_2, \quad 1 \leqslant r_2 < r_1,$$

и подставить это в новое уравнение, получив

$$r_1(x_1 + q_2 y_1) + r_2 y_1 = 1.$$

Полагая $x_2 = x_1 + q_2 y_1$, $y_2 = y_1$, приходим к уравнению $r_1 x_2 + r_2 y_2 = 1$, снова имеющему тот же вид.

(3) Продолжать процесс до тех пор, пока один из коэффициентов при x_k или при y_k не станет равным 1. Взяв решение $(1, 0)$ или $(0, 1)$ полученного уравнения, проследить проделанные шаги в обратном направлении и найти решение исходного уравнения.

Вот простой числовой пример:

$$14x + 9y = 1$$
$$14 = 9 + 5$$
$$5x + 9(x + y) = 1$$
$$5x_1 + 9y_1 = 1 \qquad x_1 = x \qquad y_1 = x + y$$
$$9 = 5 + 4$$
$$5(x_1 + y_1) + 4y_1 = 1$$

$$
\begin{array}{lll}
5x_2 + 4y_2 = 1 & x_2 = x_1 + y_1 & y_2 = y_1 \\
5 = 4 + 1 & & \\
x_2 + 4(x_2 + y_2) = 1 & & \\
x_3 + 4y_3 = 1 & x_3 = x_2 & y_3 = x_2 + y_2 \\
x_3 = 1 & y_3 = 0 & \\
x_2 = 1 & y_2 = -1 & \\
x_1 = 2 & y_1 = -1 & \\
x = 2 & y = -3 &
\end{array}
$$

Подобные алгоритмы можно придумать и для решения других диофантовых уравнений, например

$$x^2 - 2y^2 = 1.$$

Существует ли алгоритм, позволяющий выяснить, имеет или не имеет решения диофантово уравнение любой степени с любым числом неизвестных, т. е. уравнение вида

$$\sum_{i_1, i_2, \ldots, i_k = 0}^{n} a_{i_1 i_2 \ldots i_k} x_1^{i_1} x_2^{i_2} \ldots x_k^{i_k} = 0,$$

где a — целые числа? Этот вопрос до сих пор остается открытым. В 1900 г. на Международном конгрессе математиков в Париже Гильберт огласил свой знаменитый список самых важных, по его мнению, математических проблем[1]). Некоторые из них теперь уже решены. Нахождение алгоритма для диофантовых уравнений составляет десятую проблему Гильберта[2]).

Сейчас понятно, насколько эта проблема близка к проблеме разрешимости и родственным проблемам

---

[1]) См. сборник «Проблемы Гильберта» (М., «Наука», 1969).— *Прим. ред.*

[2]) В 1970 г. молодой ленинградский математик Ю. Матиясевич решил десятую проблему Гильберта, доказав отсутствие общего алгоритма для решения диофантовых уравнений (см., например, Ф. Л. Варпаховский, А. Н. Колмогоров «О решении десятой проблемы Гильберта», *Квант*, 1970, № 7, 39—44). Одновременно и независимо сходные результаты получил молодой киевский ученый Г. Чудновский. — *Прим. ред.*

логики. Поэтому не приходится удивляться, что в последние годы большие успехи на пути к ее решению были достигнуты именно благодаря методам, подсказанным математической логикой.

Алгоритм — это набор точных инструкций, описывающих, как выполнить определенную задачу. Нетрудно придумать автомат, который действовал бы по некоторому алгоритму без вмешательства человека. А существует ли универсальный автомат, который можно было бы запрограммировать на выполнение *любого* алгоритма?

Положительный ответ на этот вопрос был дан английским математиком Аланом Тьюрингом; его работа стала теоретической базой для создания современных универсальных цифровых вычислительных машин.

Универсальная машина Тьюринга по существу совсем проста. Она состоит (рис. 31) из неограниченной

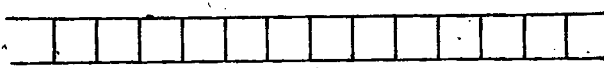


Рис. 31.

с обеих сторон ленты, разделенной на ячейки, и конечного набора символов — алфавита; для простоты можно считать, что алфавит содержит только один[1]) символ: вертикальную черточку |.

Машина Тьюринга содержит также обозреваемую ячейку, которую можно передвигать, обозревая поочередно ячейки ленты. Наконец, машина может выполнять следующие операции:

(л) сдвинуть обозреваемую ячейку на одну ячейку влево;

(п) сдвинуть обозреваемую ячейку на одну ячейку вправо;

(з) заменить символ в обозреваемой ячейке другим символом алфавита (в нашем случае — либо

---

[1]) Если придираться, то следовало бы говорить о двух символах: вторым нужно считать «пробел» или «отсутствие символа».

стереть вертикальную черточку, либо напечатать ее в пустой ячейке);

(с) остановить процедуру.

Программа представляет собой набор команд такого вида: «если в обозреваемой ячейке стоит символ ________, выполнить ________ и перейти к команде ________».

Напишем в качестве примера программу для нахождения остатка от деления целого числа на 3. Целое n изображается на ленте при помощи n вертикальных черточек, помещенных в следующих одна за другой ячейках; обозреваемая ячейка располагается справа от последней черточки. Команды нумеруются числами; «пустота» обозначается символом $*$. Команда может читаться, скажем, так: «если обозреваемая ячейка пуста, — остановиться, если нет, — сдвинуться влево и перейти к команде 2». Вот эта программа:

0	*	л	0
0	|	л	1
1	*	с	
1	|	л	2
2	*	с	
2	|	з	3
3	*	п	4
4	|	з	5
5	*	п	6
6	|	з	0

На самом деле команды можно хранить на ленте в закодированном виде (так это и делается в реальных вычислительных машинах). Однако, если бы мы здесь так сделали, наша программа стала бы намного сложнее, что вряд ли помогло бы лучше понять существо дела.

Итак, машина Тьюринга представляет собой очень простую формальную систему, которая тем не менее настолько богата, что способна воспроизводить все возможные алгоритмы. И как дополнение к результату Гёделя можно показать, что некоторые простые арифметические и комбинаторные высказывания *алгоритмически неразрешимы*. Последнее означает, что не существует программы для машины Тьюринга, позволяющей устанавливать истинность или ложность этих высказываний.

Примером неразрешимого высказывания является общая проблема тождества слов теории групп.

Допустим, что имеется четыре абстрактных символа A, B, A', B', которые можно писать один за другим, образуя произвольно длинные «слова», например

$$ABB'AAA'BBA'B.$$

Допустим также, что некоторое конечное число таких слов равно, по предположению, «единичному элементу» I группы; это означает, что когда такое слово является частью другого слова, последнее можно сократить, отбросив эту часть.

Если предположить, например, что

$$AA' = A'A = BB' = B'B = I$$

и

$$B'AA = I,$$

то слово $ABB'AAA'BBA'B$ можно сократить либо в $AABBA'B$ (применяя $BB' = AA' = I$), либо в $ABA'BBA'B$ (применяя $B'AA = I$).

Проблема состоит в следующем: существует ли алгоритм, позволяющий для любых двух слов решить, эквивалентны они или нет (т. е. представляют ли они один и тот же элемент группы)? Как показал советский математик П. С. Новиков, ответ на этот вопрос *в общем случае* отрицателен. С другой стороны, для многих частных групп (определяемых конечным числом соотношений между своими образующими) проблема тождества слов разрешима.

Тривиальным примером служит группа, определяемая соотношениями

$$AA' = A'A = BB' = B'B = I,$$
$$ABA'B' = I$$

(последнее есть не что иное, как закон коммутативности: $AB = BA$).

Для этой группы критерий эквивалентности двух слов особенно прост. Нужно сосчитать в каждом слове разность между количеством букв A и A' в нем и количеством B и B'. Два слова эквивалентны тогда и только тогда, когда отвечающие им разности одинаковы, т. е. когда две разности, образованные для первого слова, соответственно равны аналогичным разностям для второго слова.

Более трудный пример — рассмотренные ранее в этой главе группы кос. Для этих групп проблема тождества слов разрешима, и, следовательно, распознавание эквивалентности двух кос может быть предоставлено вычислительной машине (скажем, машине Тьюринга).

Вычислительные машины и их роль в математике все еще составляют предмет ожесточенных споров. Математики демонстрируют полную гамму разных отношений к этому вопросу — от равнодушия до враждебности; лишь немногие чувствуют, что вычислительным машинам предназначено сыграть важную роль в будущем развитии математики, не говоря уже об их бесспорной полезности как мощного орудия научных и технических исследований.

Идея использования механических устройств для выполнения арифметических действий и облегчения труда вычислителей очень стара. Уже древние сконструировали простые устройства и подумывали о графических методах, которые позволяли бы экономить время, выполняя простые математические действия быстрее, чем их можно было бы записывать вручную. Мы не можем удержаться от того, чтобы процитировать здесь следующий отрывок[1]: «Знаменитому и многими любимому искусству построения механических орудий положили начало Евдокс и Архит, стремившиеся сделать геометрию более привлекательной, а также с помощью чувственных осязаемых примеров

[1]) Плутарх, Сравнительные жизнеописания, т. 1, М., 1961, Марцелл, стр. 391.

разрешить те вопросы, доказательство которых посредством одних лишь рассуждений и чертежей затруднительно: такова проблема двух средних пропорциональных — необходимая составная часть многих задач, для разрешения которой оба применили механическое приспособление, строя искомые линии на основе дуг и сегментов. Но так как Платон негодовал, упрекая их в том, что они губят достоинство геометрии, которое от бестелесного и умопостигаемого опускается до чувственного и вновь сопрягается с телами, требующими для своего изготовления длительного и тяжелого труда ремесленника, — механика полностью отделилась от геометрии и, сделавшись одною из военных наук, долгое время вовсе не привлекала внимания философии.»

Паскаль сконструировал машину для выполнения арифметических операций; Лейбниц думал о логической машине, которая в конце концов могла бы решать все проблемы математических наук. Высказав идею формальной системы и мысль о возможности оперировать в ней механически, Лейбниц предвосхитил многие недавние достижения логики. Уже в 19-м веке Бэббидж и другие разработали механические вычислительные устройства. Однако только в самое недавнее время (примерно после 1940 г.) благодаря новым средствам электронной техники разработка вычислительных машин достигла таких масштабов, о которых раньше нечего было и думать. Это обещает не только сильно увеличить объем математических сведений, но и повлиять на ход и направления самих математических исследований.

На протяжении всей истории математики новые идеи рождались и как проблески интуиции, и как результат терпеливых наблюдений. Изучение и обсуждение накопленных фактов подсказывали новые обобщения. В теории чисел, например, многие свойства целых чисел были впервые обнаружены путем экспериментирования с «малыми» числами. Сам Гаусс, когда его спросили, как он пришел к некоторым своим общим идеям, ответил: «Durch planmässiges Tattonieren» — путем планомерного экспериментирования «на пальцах»[1]). Трудно переоценить наводящую роль примеров и указаний, содержащихся в частных случаях; они в большой мере определяют направления, которых придерживается математик в своих исследованиях.

Сейчас машины выполняют арифметические операции сложения, вычитания, умножения и деления двух чисел с точностью 10^{-12} быстрее, чем за одну миллионную долю секунды. Они снабжены памятью (т. е. запоминающими устройствами), где хранятся сотни тысяч чисел, которые можно быстро оттуда извлекать. Время введения или извлечения данных из памяти также составляет величину порядка одной миллисекунды. Эти устройства могут выполнять простые логические операции (булевы операции); они снабжены также многими «командами», автоматически выполняющими простые комбинаторные действия.

Конструирование и сооружение вычислительных машин, а также разработка методов представления математических задач в эффективном и точном виде особенно ускорились в связи с техническими проблемами, поставленными второй мировой войной. Работа по использованию электронных вычислительных машин в проблемах чистой науки непрерывно расширялась, захватывая все новые области и проникая в глубь ранее освоенных территорий. Основное место здесь занимали исследования математических вопросов теоретической физики. В самой же математике применение электронных машин затрагивало главным образом комбинаторный анализ и теорию чисел.

Могло бы показаться, что по-настоящему интересные вопросы требуют обращения с очень большими числами, недоступными никакой машине, сколь бы велик ни был объем ее памяти. Однако во многих случаях предельное или асимптотическое поведение функций достаточно хорошо прослеживается уже на

[1]) Укажем, например, что в поисках общих закономерностей, относящихся к периодическим десятичным дробям, Гаусс не остановился перед столь трудоемкой работой, как выписывание периодов, получающихся при обращении в десятичные дроби всех простых дробей со знаменателями до 1000. — *Прим. ред.*

малых или средних значениях аргументов. Например, изучение плотности распределения простых чисел вплоть до 10 000 дает вполне удовлетворительное представление об асимптотическом ее значении. Мы уже упоминали теорему о распределении простых чисел, утверждающую, что число $\pi(n)$ простых чисел от 1 до n асимптотически равно $n/\log n$. Это, грубо говоря, означает, что для больших n в последовательности натуральных чисел встречается примерно одно простое число между n и $n+[\log n]$, где $[\log n]$ — наибольшее целое, не превосходящее $\log n$. Иногда в этом интервале будет в точности одно простое число, иногда два, иногда три, а иногда и ни одного. Представляется разумным разбить целые числа на классы, обозначив через C_0 класс всех целых n, для которых в указанном интервале нет ни одного простого числа, через C_1 — класс всех целых чисел, для которых в этом интервале имеется ровно одно простое число, через C_2 — класс тех n, для которых в этом интервале имеется точно два простых числа, и т. д. Сейчас теория чисел, по-видимому, не располагает средствами, которые позволили бы доказать, что для этих классов существуют асимптотические плотности, т. е. доказать, например, что существует предел $\gamma_0 = \lim\limits_{N\to\infty} (\gamma_0(N)/N)$, где $\gamma_0(N)$ — число целых чисел между 1 и N, принадлежащих классу C_0. Для современной вычислительной машины, способной запомнить все простые числа вплоть до 100 000 000, исследование поведения таких отношений не составляет особого труда. По-видимому, они стремятся к определенным пределам, первый из которых $\gamma_0 \approx 0{,}30\ldots$, а другие таковы: $\gamma_1 \approx 0{,}42\ldots$; $\gamma_2 \approx 0{,}21\ldots$; $\gamma_3 \approx 0{,}05\ldots$; $\gamma_4 \approx 0{,}006\ldots$ и т. д. И вот простое наблюдение подсказывает теоремы, еще не доказанные: во-первых, что рассматриваемые пределы существуют, и, во-вторых, что они стремятся к нулю с возрастанием номера класса.

В комбинаторном анализе (т. е. анализе конечных конфигураций и проблем, связанных с возможной эволюцией этих конфигураций) такая эвристическая работа уже принесла большую пользу. В частности, недавно при помощи примера, подсказанного численным расчетом, выполненным на машинах, была опровергнута гипотеза Эйлера о взаимно ортогональных латинских квадратах.

«Грубая сила» вычислительной машины позволяет изучать сложные вопросы перечисления вариантов, возникающие в комбинаторном анализе. Как и в приведенном выше примере из теории чисел, машина может перебрать все возможности при малых значениях переменной n и изучить зависимость от n, помогая тем самым сформулировать гипотезу об асимптотическом поведении. Приведём такой пример. Пусть задано множество E целых чисел от 1 до n и *случайно* выбраны две подстановки S_1 и S_2 элементов этого множества. Каково среднее число элементов группы, порожденной этими подстановками? Чтобы решать такие задачи на вычислительной машине, можно использовать процедуру статистических выборок, известную под названием метода Монте-Карло. Этот подход основан на чрезвычайно простой идее, но если бы не вычислительные машины, от него было бы мало проку.

Вот как он работает. Пусть $g(S_1, S_2)$ — число элементов в группе, порожденной подстановками S_1 и S_2 (т. е. число *различных* подстановок, которые можно получить, образуя всевозможные произведения вида $S_1^{k_1}S_2^{l_1}S_1^{k_2}S_2^{l_2}\ldots$). Искомое среднее равно

$$\frac{\sum g(S_1, S_2)}{\text{число различных пар } (S_1, S_2)},$$

где $\sum$ в числителе означает суммирование по всем различным парам (S_1, S_2). Знаменатель можно сосчитать; он равен

$$\frac{(n!)^2 + n!}{2}.$$

Уже при $n = 5$ эта формула дает число 7260, и, таким образом, даже при таком малом n нужно перебрать 7260 групп и найти их порядки. Эта задача напоминает задачу определения среднего веса новорожденных мальчиков. Вместо того чтобы усреднять веса *всех* новорожденных мальчиков, образуют (представительную) *случайную выборку* и по ней производят усреднение. Как образовать ее, избежав нежелательных смещений разного рода, — это уже часть искусства и науки статистики. Машину можно запрограммировать на то, чтобы она образовывала случайные выборки из множества всех пар подстановок. Такой метод позволяет получить очень хорошее приближение искомого среднего.

Однако развитие быстродействующих электронных машин открывает еще более широкие перспективы. Кроме общих возможностей разнообразного экспериментирования в задачах чистой математики или разработки предварительных идей, касающихся физических теорий, можно представить себе машины, действительно работающие в формальных системах математики. Грубо говоря, существующие машины действуют на некотором множестве заданных команд (блок-схема и код), которые задаются раз и навсегда и заставляют машину автоматически выполнять решение численных или комбинаторных задач. Ход работы машины полностью предписан. Единственное, в чем она может проявить некоторую «гибкость», — выбрать тот или иной ход вычисления в зависимости от полученных числовых значений. До сего времени так называемые решения, принимаемые машиной, практически сводились лишь к ограниченному набору изменений логического хода вычисления.

Но можно представить себе ситуацию, когда машина устроена так, что удается поддерживать постоянную связь между ней и разумным оператором, который в ходе вычислений (в зависимости от своих наблюдений и истолкования полученных результатов) может изменять логическую структуру самой задачи. Это намного расширило бы область использования машин. Они могли бы не только принять на себя все тяготы элементарных алгебраических или аналитических выкладок, но и помочь исследователю-математику, скажем, в поисках примеров или контрпримеров, быстро показывая ему визуально на экране то, что его интересует, и тем самым направляя или проверяя его интуицию.

Такую помощь машина может оказать, например, при изучении свойств функций нескольких переменных и при исследовании преобразований многомерных пространств.

Во многих задачах важно найти экстремальные значения функции нескольких действительных переменных $f(x_1, x_2, \ldots, x_n)$, заданной аналитически, т. е. выраженной через элементарные функции. Известно, что обычная процедура поисков численных значений локальных минимумов или максимумов одной функции отнимает очень много времени. Если независимых переменных много, скажем 5 или больше, не существует никакого по-настоящему эффективного метода нахождения всех критических точек. Допустим теперь, что можно быстро вычислять значения этой функции в точках некоторой сетки на плоскости, образующей сечение многомерного пространства, являющегося областью задания функции. Полученный таким способом «график» функции двух переменных (поверхность) проектируется на электронно-лучевой видеоэкран. (Программа вычисления аксонометрических проекций имеется в готовом виде.) Тогда можно быстро обнаружить область или несколько областей, где функция может принимать минимальные значения. Путем быстрого изменения масштаба и увеличения (т. е. подразделения таких «подозреваемых на минимум» областей на большее число точек) вычислительная машина может действовать как сколь угодно сильный микроскоп. Таким образом, вместо слепых рецептов, входящих в программу поиска критических точек, вступит в действие зрительное восприятие человека, которое «срабатывает» все еще намного быстрее, чем любая известная программа автоматического распознавания.

При изучении функций трех или четырех переменных также хотелось бы иметь возможность заставлять машину быстро выбирать требуемое двумерное сечение, устанавливать на нем сетку независимых узловых точек, вычислять значения функций в этих точках и давать их перспективное изображение на экране. Не менее полезной оказалась бы способность машины выполнять замену независимых или зависимых переменных при помощи общего линейного преобразования.

Еще большим шагом вперед была бы возможность проводить при помощи машинных вычислений серию «уроков», в результате которых оператор после некоторой практики научился бы «чувствовать» четырехмерное пространство.

Рассмотрим в трехмерном пространстве задачу «продевания» некоторого тела сквозь замкнутую пространственную кривую — упражнение на метод проб и ошибок. Чтобы решить вопрос о том, можно ли протолкнуть данное тело сквозь данную кривую, никаких простых критериев, связанных с проекциями, по-видимому, недостаточно. Физический процесс такого продевания можно имитировать при помощи машины, заставив ее вычислять последовательные положения тела в соответствии с данными от руки указаниями (быстро передающимися в виде чисел) о поворотах и переносах в трехмерном пространстве и проверять после каждого пробного смещения, соприкасаются или нет рассматриваемые множества.

Уже созданы программы, позволяющие машине оперировать с символическими выражениями и отыскивать способы обращения с некоторыми простыми системами аксиом для получения формальных доказательств и поиска теорем. Эта работа только начинается. Ясно, что можно формально оперировать с многочленами и рациональными дробями, если запрограммировать алгебраические операции над ними. Существуют программы, заставляющие машину выполнять формальное дифференцирование и находить неопределенные интегралы широкого класса элементарных функций. Имеются также интересные программы доказательства теорем в системах евклидовой или проективной геометрий. Машины уже выдали любопытные доказательства свойств треугольников и т. п., иногда не совпадающие с принятыми в школьных курсах.

Что касается работы, связанной с логикой, то до сих пор машины владели лишь элементарными булевыми операциями, или, что равносильно, исчислением высказываний. Следующим шагом было бы научить их оперировать с кванторами: «существует x, такой, что ...» и «для всех x ...»; это значительно расширило бы их логические возможности.

Большой интерес представляет тот факт, что вычислительные машины оказались источником новых захватывающих математических проблем.

Упомянем в качестве примера класс задач, связанных с так называемой *алгоритмической сложностью.*

Как-то всегда кажется, что умножение — более сложная операция, чем сложение. Существует ли способ перевести это смутное ощущение в точное математическое утверждение?

Разумную формулировку подсказывает обращение к машине Тьюринга.

Сложение двух n-разрядных чисел можно выполнить на машине Тьюринга за число шагов, примерно пропорциональное n. Точнее, отношение этого числа шагов к n стремится к конечному пределу, когда n стремится к бесконечности. Пусть $M(n)$ — минимальное число шагов, нужных машине Тьюринга для перемножения двух n-разрядных чисел. Если бы удалось доказать, что

$$\lim_{n\to\infty} \frac{M(n)}{n} = \infty,$$

то это соотношение можно было бы истолковать как означающее, что умножение и в самом деле есть более сложная операция, чем сложение. Однако вместо этого было установлено, что для любого $\varepsilon > 0$

$$\lim_{n\to\infty} \frac{M(n)}{n^{1+\varepsilon}} = 0;$$

отсюда видно, что умножение по своей сложности не так уж сильно отличается от сложения.

Новые проблемы требуют для своего решения новых методов, выявляют новые связи и подкрепляют старые. С этой точки зрения вычислительные машины уже внесли заметный вклад как в *проблематику*, так и в *методологию* математики. Не приходится сомневаться, что их влияние будет постоянно усиливаться, несмотря на негодование некоторых современных последователей Платона.

Электронные вычислительные машины оказались неоценимым инструментом исследования и для других наук. Одно только перечисление отдельных задач математической физики, решения которых, полученные при помощи машин, помогли пересмотреть существующие теории и подсказали новые свойства сложных физических систем, заняло бы сотни, а быть может, и тысячи страниц. Особенно важная роль предназначена машинам в новых быстро развивающихся областях биологии. С их помощью уже удалось расшифровать структуру некоторых органических молекул (в частности структуру белка миоглобина). Математически такого рода задача состоит в определении пространственного расположения отдельных атомов по дифракционным картинам, которые дает молекула в целом. При этом используются методы обращения преобразований Фурье и манипуляции с большими массивами статистических данных. Можно без преувеличения сказать, что такие вопросы нельзя было бы решать, не пользуясь современными вычислительными средствами.

В заключение следует хотя бы упомянуть об одном важном обстоятельстве, которое часто остается незамеченным: как вычислительные машины заставили нас по-новому взглянуть на целый ряд разнообразных проблем.

Рассмотрим, например, вопрос о машинном переводе с одного языка на другой. Современные вычислительные машины вполне приспособлены для запоминания обширных словарей; они позволяют извлекать требуемые данные с фантастической скоростью. Однако этого недостаточно, если не дать машине хотя бы самых скромных сведений о правилах грамматики и синтаксиса. Задача обучения автомата требует резкого критического пересмотра собственных лингвистических знаний программиста. Он не может рассчитывать на то, что машина обладает необходимым запасом культуры и теми сложными и тонкими психологическими особенностями, на которые он, вероятно, бессознательно опирался бы, имея дело с человеком. Проблема, которая встает перед ним, чрезвычайно глубока: по существу это вопрос о том, сколь многое из того, что мы относим к «интеллекту», можно заменить большим объемом памяти и скоростью извлечения из нее знаний.

Можно думать, что мы близки к точной формулировке таких проблем и что вычислительные машины повлияли на наши взгляды на подобные вопросы и даже, вероятно, изменили их.

Мы неоднократно, хотя и не всегда явно, подчеркивали, что своей уникальностью математика обязана приверженностью к аксиоматическому методу.

Как мы говорили, этот метод состоит в том, что исходя из нескольких предложений (аксиом), истинность которых считается не требующей доказательства, выводят другие предложения, применяя *только* правила логики.

Аксиомы предназначены для описания простых свойств рассматриваемых объектов в надежде на то, что эти свойства полностью отразят существенные черты этих объектов. Однако как узнать, действительно ли система аксиом выделяет именно то, что было задумано?

Система аксиом не выполняет эту задачу (установить это несколько легче, чем противоположное), если к ней можно добавить новое относящееся к тем же объектам предложение A или его отрицание $\bar{A}$ («не A») и в обоих случаях получить непротиворечивую систему. Иными словами, система аксиом *не является полной* (*неоднозначно* характеризует объекты), если существует предложение A, *независимое* по отношению к аксиомам этой системы (т. е. такое, что присоединение к системе как утверждения A, так и его отрицания $\bar{A}$ не приводит к противоречию).

Вопрос о независимости аксиом восходит к древности. Попытка вывести аксиому о параллельных (пятый постулат Евклида, утверждающий, что через точку, не лежащую на прямой l, можно провести одну и только одну прямую, параллельную l) из остальных аксиом оказалась одним из самых мощных двигателей развития математики после Евклида. То, что пятый постулат не удавалось ни доказать, ни опровергнуть, несмотря на настойчивые усилия многих поколений математиков, несомненно, способствовало возведению евклидовой геометрии в некий абсолют. Кант, например, полагал, что евклидова геометрия доставляет единственно возможный способ дедуктивной трактовки свойств пространства; поэтому иногда его обвиняют в том, что он задержал открытие неевклидовых геометрий.

Это сражение с пятым постулатом носило почти трагический характер. Вот, например, отрывок из письма Бойаи-старшего, в котором он умолял сына прекратить свои исследования: «Я обследовал все рифы этого адского Мертвого Моря и каждый раз возвращался с поломанной мачтой и разорванными парусами».

Даже после того как Бойаи-младший и Лобачевский независимо пришли к открытию первых неевклидовых геометрий, логический статус пятого постулата оставался неясным.

Полностью дело прояснилось несколько позднее. Сначала Бельтрами открыл, что геометрия на определенной поверхности (названной псевдосферой, поскольку это поверхность *постоянной отрицательной кривизны*) служит реализацией геометрии Бойаи — Лобачевского, а затем Клейн и Пуанкаре построили совсем простые плоские модели, в которых реализовалась эта новая геометрия.

Все это привело к возникновению общего метода доказательства независимости аксиом. Коротко говоря, такой метод состоит в следующем. Для того чтобы доказать, что аксиома A является независимой по отношению к некоторой системе аксиом S, к системе S присоединяют $\bar{A}$ — отрицание утверждения A — и пытаются найти объекты, удовлетворяющие S и $\bar{A}$. Если это можно сделать, то система, состоящая из S и $\bar{A}$, непротиворечива; тогда A является независимой по отношению к системе S и система S может быть пополнена аксиомой A.

Обычно при построении объектов, образующих *модель* для некоторой системы аксиом, приходится использовать другую систему, непротиворечивость которой считается уже установленной или не требующей доказательства. Так, например, аналитическая геометрия на плоскости является моделью планиметрии: точки интерпретируются как упорядоченные пары

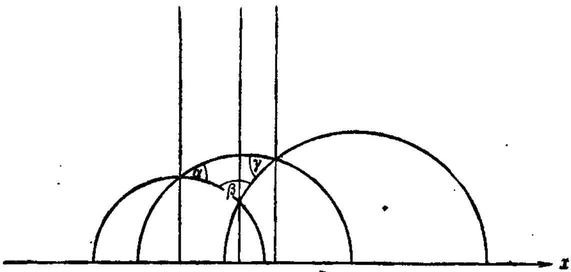


Рис. 32.

(x, y) действительных чисел, прямые — как множества пар (x, y), удовлетворяющих линейному уравнению вида $ax + by + c = 0$, и т. д. Такая интерпретация опирается на систему действительных чисел, а ведь каждый математик, несмотря на определенные логические трудности, все-таки исходит из предположения, что алгебра действительных чисел непротиворечива.

Вот как доказывается независимость пятого постулата в планиметрии. Рассмотрим верхнюю полуплоскость обычной плоскости, исключая из нее ось x (рис. 32). «Точками» новой геометрии будем считать точки этой полуплоскости, а «прямыми» — полуокружности с центрами на оси x (они пересекают ее под прямым углом) и (обычные) прямые, перпендикулярные оси x. «Углы» между «прямыми» определим обычным образом (т. е. как *евклидовы* углы между касательными к изображающим «прямые» окружностям). Чтобы завершить описание этой геометрии, осталось дать определение конгруэнтности, или, что равносильно, ввести «жесткие движения». Такими движениями будем считать взаимно однозначные преобразования рассматриваемой полуплоскости в себя, переводящие «прямые» (т. е. полуокружности с центрами на оси x или прямые, перпендикулярные оси x) в «прямые» и не искажающие углов (требование «жесткости»).

Выбрав на оси x начало координат и сопоставив каждой точке P верхней полуплоскости комплексное число z:

$$P \to z = x + iy, \quad y > 0,$$

можно показать, что жесткими движениями являются преобразования $T(z)$ вида

$$T(z) = \frac{\alpha z + \beta}{\gamma z + \delta},$$

где α, β, γ, δ — действительные числа, причем $\alpha\delta - \beta\gamma > 0$.

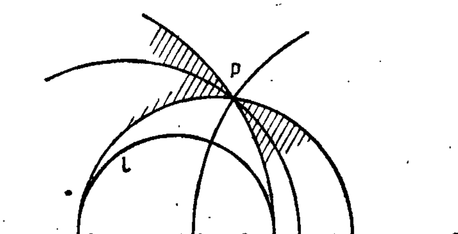


Рис. 33.

Понятие «между» вводится очевидным образом. После этого можно проверить, что выполняются *все* гильбертовы аксиомы евклидовой планиметрии (в том числе и несколько туманная аксиома непрерывности[1])), *кроме* аксиомы о параллельных.

Как видно из рис. 33, прямые, проходящие через точку P и параллельные прямой l, заполняют весь

[1]) Как упоминалось выше, аксиома непрерывности устанавливает сохраняющее порядок взаимно однозначное соответствие между точками прямой и действительными числами. Для нашей конкретной модели такое соответствие легко указать в явном виде.

заштрихованный на этом рисунке *угол*. Может вызвать удивление тот факт, что результат многовековых усилий выражается (и притом вполне точно) в столь немногих словах. Бойаи и Лобачевскому пришлось гораздо труднее, ибо в те времена еще не существовало понятия модели. Они исходили непосредственно из аксиом и путем обычных евклидовых процедур вывода получали заключения, которые наверняка казались им чрезвычайно странными; с железной выдержкой они должны были подчинить свои привычные интуитивные представления не допускающим отклонений требованиям логики.

Нигде в математике вопросы полноты и независимости аксиом не играют такой большой роли, как в теории множеств. Поскольку теория множеств единодушно признается фундаментом всей математики, ее аксиоматические основания — вопрос первостепенной важности.

В то же время в теории множеств попадаются предложения вроде аксиомы выбора, которая имеет такие странные следствия, что некоторые математики, хотя и не отвергают ее открыто, все же предпочитают ее избегать.

Имеется также знаменитая канторовская гипотеза континуума, доказательство (или опровержение) которой Гильберт считал настолько важным, что в своем списке нерешенных проблем поставил ее первой.

Через несколько лет после опубликования своей работы о неразрешимых предложениях Гёдель показал, что и аксиому выбора и гипотезу континуума можно рассматривать как истинные в таких формальных системах теории множеств, как система Френкеля или система Гильберта — Бернайса. Иначе говоря, он показал, что эти предложения либо доказуемы в такой системе, либо являются независимыми по отношению к остальным аксиомам и могут при желании быть добавлены к этой системе.

Полное выяснение статуса аксиомы выбора и гипотезы континуума было достигнуто в 1960 г., когда Поль Коэн показал, что в обычных формальных системах теории множеств аксиому выбора и гипотезу континуума можно, не впадая в противоречие, считать *не выполняющимися*!

Как сказано выше, результат Гёделя состоял в том, что ни аксиома выбора, ни гипотеза континуума не являются *доказуемо* ложными. Этот результат был получен путем построения *модели* теории множеств (удовлетворяющей, например, аксиомам Френкеля или Гильберта — Бернайса), в которой истинно какое-нибудь одно из этих предложений или они оба. Существование такой модели не означает, что рассматриваемое предложение (скажем, гипотеза континуума) выводится из аксиом, точно так же, как построения модели геометрии, удовлетворяющей всем аксиомам Евклида включая пятый постулат, недостаточно для вывода этого постулата из остальных аксиом. Что же касается Коэна, то он построил *модели* теории множеств, в которых аксиома выбора не выполняется, и другие модели, в которых она выполняется, но не выполняется гипотеза континуума. Он указал систему, удовлетворяющую всем обычным аксиомам теории множеств включая аксиому выбора, в которой, однако, континуум имеет «очень большую» мощность, и, таким образом, в этой системе существуют множества промежуточной мощности между $\aleph_0$ (мощностью множества всех целых чисел) и $\mathfrak{c}$ — мощностью континуума[1]).

Эти результаты следует рассматривать как определенно разрешающие проблему Кантора, по крайней мере в рамках существующих формализаций теории множеств. На наш взгляд, они бросают новый вызов и открывают новые перспективы исследования оснований математики. Следует, однако, помнить, что мы много раз употребляли слова «формальные аксиоматические системы» теории множеств. Возникает вопрос: действительно ли эти системы являются достаточно общими, чтобы охватить всю нашу интуицию, которая в некотором смысле более существенна, чем

---

[1]) По поводу этих результатов см., например, П. Дж. Коэн и Р. Херш, Некантороввская теория множеств, *Природа*, № 4 (1969), 43—55. — *Прим. ред.*

любое формализованное множество написанных на бумаге выражений, выдаваемых за отражение этой интуиции? *При современном состоянии* математического мышления такой вопрос следует отнести к метаматематике. Однако он имеет важное философское значение; как мы видели на предыдущих примерах, в процессе развития науки неоднократно идеи, которые первоначально были метаматематическими, становились в конце концов частью самой математики. Таким образом, метаматематика представляет значительный потенциальный математический интерес.

Скорее всего, в будущем так и не удастся прийти ни к какой определенной конечной системе аксиом, рассматриваемой как окончательная. Напротив, подобно тому, как это происходит в мире живого, будут постепенно появляться все новые аксиомы. Возможно, система аксиом будет развиваться, так сказать, «генетическим» путем на основе существующих аксиом в результате согласованной работы математиков над их следствиями. Можно было бы еще добавить, что никакая из рассматривавшихся до сих пор формальных систем не дает адекватного воплощения того представления о бесконечном, которого бессознательно придерживаются математики; можно даже отважиться на гипотезу, что такая *формальная* система вообще невозможна.

Обсуждая темы и тенденции математики, мы уделили много внимания основаниям, особенно основаниям теории множеств. Однако мы не хотели создать впечатление, будто глубоко прояснившие дело работы Гёделя и Коэна непосредственно затронули главное русло математики. На самом деле бóльшую часть математики они не задели вовсе. С другой стороны, прояснение оснований геометрии, принесенное открытием неевклидовых геометрий, оказало глубокое влияние на математику, физику и астрономию.

Возможно, это произошло потому, что теоретико-множественные основания математики слишком всеобъемлющи и носят слишком общий характер, чтобы играть жизненно важную роль в конкретных задачах повседневной математики, а, может быть, потому, что они имеют дело скорее с самим процессом дедукции, чем с его плодами.

Как бы то ни было, работа над основаниями *всей математики в целом* привела к отрицательному результату, ибо она выявила слабые стороны аксиоматического метода. В теории множеств она породила серьезные сомнения в существовании формальных систем, способных дать такое описание, которое отвечало бы представлению математика о множествах.

Напротив, работа над основаниями геометрии имела другой, значительно более конструктивный результат. Оказалось, что очень многое из наших интуитивных представлений о пространстве может быть формализовано в виде дедуктивной системы. Более того, сама эта формализация содержит в себе зародыши новых миров, подобных открытым Лобачевским и Бойаи, миров, которые сначала противоречили физической интуиции, но лишь затем, чтобы слиться с ней позднее в рамках теории относительности. Наконец, эта работа оказала важное влияние на дифференциальную геометрию — один из центральных разделов математики, который все еще активно изучается. Трудно избежать искушения и не сделать из всего этого вывод, что существует какое-то неопределяемое глубокое различие между проблемой аксиоматизации отдельной ветви математики, обязанной своим происхождением *внешним* стимулам, и проблемой аксиоматизации *внутренних* процессов мышления.

ГЛАВА 3

СВЯЗЬ С ДРУГИМИ НАУКАМИ

Вопрос о связи между математикой и естественнонаучными дисциплинами веками ставил в затруднение философов и историков науки.

Вряд ли стоит сомневаться в том, что источником многих математических понятий и теорий послужил «внешний мир». Но однажды постигнутые, эти понятия и теории начинали развиваться совершенно независимо. Они поднимались к высотам абстракции, освобождаясь от пут своего конкретного (даже «низменного») происхождения. В процессе этой эволюции чисто интроспективным путем рождались новые понятия и теории, которые в свою очередь чудодейственным образом оказывали решающее влияние на ход научного прогресса уже за пределами собственно математики.

Рассмотрим в качестве примера геометрию. Она возникла из опыта древних землемеров и астрономов и на своем первом великом этапе развития достигла кульминации в «Элементах» Евклида, которые веками служили непреложным образцом логической строгости и совершенства.

Оторвавшись от внешнего мира, из которого она возникла, геометрия продолжала развиваться, питаясь своими собственными проблемами. Среди них была и проблема пятого постулата — столь же неуловимая, сколь и привлекательная.

Как мы видели в гл. 2, задача была чисто логической: можно ли вывести указанную аксиому (постулат) из остальных аксиом евклидовой геометрии.

Бойаи и Лобачевский первыми дали отрицательный ответ на этот вопрос, построив систему геометрических предложений (включающую отрицание пятого постулата), которые находились в таком взаимно однозначном соответствии с их евклидовыми аналогами, что противоречие в одной из этих систем немедленно повлекло бы за собой противоречие в другой.

Интересно отметить, что ни Бойаи, ни Лобачевский не имели отчетливого ощущения «реальности» своей геометрии. Лобачевский называл ее «мнимой», а Бойаи взволнованно писал отцу: «... из ничего я создал новый и удивительный мир».

Лишь много лет спустя геометрия Бойаи — Лобачевского помогла Риману найти глубокий и открывающий новые перспективы подход к неевклидовым геометриям. Созданный в результате математический аппарат был положен в основание общей теории относительности Эйнштейна.

Этот пример с геометрией — вероятно, самая драматическая, но далеко не единственная иллюстрация превращений, которые претерпевают математические понятия и идеи.

Исследуя, как остывает Земля, Фурье пришел к проблеме представления периодической функции в виде ряда, состоящего из синусов и косинусов:

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

Та же проблема возникает при попытке разложить сложное периодическое колебание (например, звуковую волну, создаваемую музыкальным инструментом) на простые «чистые тоны» (синусоидальные колебания).

Эти задачи физики дали мощный толчок изучению рядов, подобных написанному выше, что привело к созданию чисто математической теории тригонометрических рядов.

По мере развития этой теории стало очевидно, что некоторые ее разделы совершенно не связаны с синусоидальностью «чистых тонов». В действительности большинство результатов останутся верными, если появившиеся из физических соображений синусы и косинусы заменить функциями $\varphi_n(x)$, подчиненными единственному условию

$$\int_0^1 \varphi_n(x)\varphi_m(x)\,dx = \begin{cases} 0, & m \neq n, \\ 1, & m = n. \end{cases}$$

Это условие является аналогом условия взаимной перпендикулярности векторов евклидова пространства (см. § 15 гл. 1) и требования, чтобы эти векторы имели единичную длину. Итак, задача представления функции в виде ряда

$$\sum_{n=1}^{\infty} c_n \varphi_n(x)$$

стала аналогом задачи разложения вектора на взаимно перпендикулярные компоненты.

Эта и другие аналогии того же рода привели к появлению понятия простейшего бесконечномерного пространства — так называемого гильбертова пространства. И снова чудо: гильбертово пространство оказалось подходящим «математическим каркасом» квантовой механики.

Известно, что развитие математики, особенно в некоторые периоды, в значительной мере определялось задачами физики и астрономии.

Так, исчисление бесконечно малых — самый крупный, по-видимому, шаг на пути эволюции математических понятий и методов — было развито Ньютоном для решения задач динамики и, в частности, задач, возникающих при изучении движения планет. Коронным достижением Ньютона был вывод законов Кеплера[1]) из закона всемирного тяготения.

Закон всемирного тяготения утверждает, что два тела притягиваются с силой, прямо пропорциональной их массам и обратно пропорциональной квадрату рас-

---

[1]) Законы Кеплера таковы: 1) планеты движутся по эллипсам, в одном из фокусов которых находится Солнце; 2) отрезок (радиус-вектор), соединяющий Солнце и планету, заметает за равные промежутки времени равные площади (поэтому каждая планета движется быстрее, когда она ближе к Солнцу); 3) *квадраты* периодов обращения планет вокруг Солнца пропорциональны *кубам* их расстояний от Солнца.

стояния между ними. Постулированный Ньютоном второй закон механики (сила пропорциональна массе, умноженной на ускорение) позволяет установить, что ускорение планеты обратно пропорционально квадрату ее расстояния от Солнца и направлено к Солнцу вдоль отрезка, соединяющего ее с Солнцем. Поскольку ускорение есть вторая производная радиуса-вектора планеты, предыдущее заключение приводит к уравнению, связывающему вторую производную вектора с самим вектором. Такое уравнение называется дифференциальным, так как в него входит производная искомой функции от времени. Ньютон вывел и решил это уравнение, получив в качестве следствия все три закона Кеплера.

Трудно передать, какое громадное воздействие оказало это великое деяние Ньютона на развитие науки. Оно, несомненно, положило начало теоретической физике и дало образец использования математических понятий и представлений для описания физических явлений.

Основных операций исчисления бесконечно малых — дифференцирования и интегрирования — оказалось вполне достаточно для того, чтобы сформулировать все физические законы, открытые в 18-м и 19-м веках. Теория упругости, гидродинамика, термодинамика и великое достижение Максвелла — теория электромагнитного поля — все это дань почти непостижимой многосторонности этого исчисления. Не удивительно поэтому, что анализ — раздел математики, выросший на почве дифференциального и интегрального исчисления, — стал поистине языком точных наук и превратил математиков в полноправных участников битв за овладение тайнами природы.

В течение двух прошлых столетий физика становилась все более математической, математика же, с одной стороны, все сильнее проникала в физику, а с другой, все больше проникалась физическим духом. Многие крупные математики того времени были и ведущими физиками. Традиция тесного сотрудничества между двумя этими науками продолжается и до наших дней, хотя его масштабы сильно сократились.

О том, сколь плодотворным и многообещающим являлось такое сотрудничество, свидетельствует, например, предсказание электромагнитных волн и создание электромагнитиой теории света. К середине 19-го века накопилось много экспериментальных данных, касающихся электромагнитных явлений. На базе этих даниых, сочетая строгую дедукцию с дерзким предвидеиием, Максвелл сумел прийти к системе дифференциальных уравнений, вобравших в себя все, что было известно в то время об электричестве и магнетизме.

Особенно простой вид уравнения Максвелла принимают для электромагнитного поля в вакууме — они содержат лишь две векторные величины: напряженность электрического поля $\vec{E}$ и напряженность магнитного поля $\vec{B}$. Вот эти уравнения:

$$\nabla\vec{E}=0, \qquad \nabla\vec{B}=0,$$

$$\alpha\frac{\partial\vec{E}}{\partial t}=\nabla\times\vec{B}, \qquad \beta\frac{\partial\vec{B}}{\partial t}=-\nabla\times\vec{E}.$$

Коэффициенты α и β зависят от выбора едииц. Из этих уравнений путем несложных *математических* выкладок можно вывести, что напряженность электрического поля $\vec{E}$ удовлетворяет уравиению

$$\frac{\partial^2\vec{E}}{\partial t^2}=\frac{1}{\alpha\beta}\nabla\cdot\nabla\vec{E}=\frac{1}{\alpha\beta}\left(\frac{\partial^2\vec{E}}{\partial x^2}+\frac{\partial^2\vec{E}}{\partial y^2}+\frac{\partial^2\vec{E}}{\partial z^2}\right);$$

то же самое уравнение получается и для напряжениости магнитного поля $\vec{B}$:

$$\frac{\partial^2\vec{B}}{\partial t^2}=\frac{1}{\alpha\beta}\nabla\cdot\nabla\vec{B}=\frac{1}{\alpha\beta}\left(\frac{\partial^2\vec{B}}{\partial x^2}+\frac{\partial^2\vec{B}}{\partial y^2}+\frac{\partial^2\vec{B}}{\partial z^2}\right).$$

Величина $1/\alpha\beta$ имеет размерность квадрата скорости; она может быть определена экспериментально. Оказывается, что

$$1/(\alpha\beta)=c^2,$$

где c — скорость света!

Незадолго до Максвелла стало известно, что локальное возмущение в изотропной упругой среде (находившейся в состоянии покоя в начальный момент времени) распространяется в виде волн, причем это распространение описывается *волновым уравнением*

$$\frac{\partial^2U}{\partial t^2}=c^2\left(\frac{\partial^2U}{\partial x^2}+\frac{\partial^2U}{\partial y^2}+\frac{\partial^2U}{\partial z^2}\right),$$

где $U(x, y, z, t)$ — отклонение от начального положения покоя в точке (x, y, z) в момент времени t. Здесь постоянная c — это скорость распространения волн в рассматриваемой среде.

Максвелл был поражен тем фактом, что электрический и магнитный векторы подчиняются волновому уравнению, и пришел к выводу, что электромагнитные возмущения тоже распространяются в виде волн. Это великолепное теоретическое предсказание блестяще подтвердилось в 1886 г., когда Генрих Герц экспериментально получил электромагнитные волны. Поскольку электромагнитные волны распространяются со скоростью света, Максвелл предположил, что свет является одной из форм электромагнитного излучения. Это предположение также полностью подтвердилось многочисленными экспериментами и дальнейшими теоретическими выводами. В результате было достигнуто более глубокое понимание природы света.

Пример максвелловской теории электромагнитного поля иллюстрирует и другое (в некотором смысле более тонкое) взаимодействие математических и физических идей. Оно связано с тем, что уравнения Максвелла, в отличие от законов Ньютона, *не инвариантны* относительно преобразований Галилея (см. § 16 гл. 1).

С другой стороны, эти уравнения сохраняют свой вид при преобразованиях Лоренца (§ 16 гл. 1). Этот чисто математический факт следует из *формы* уравнений Максвелла, и в принципе его можно было бы обнаружить, не имея ни малейшего представления о физическом содержании этих уравнений. Однако дерзкое требование изменить законы динамики так, чтобы они тоже стали инвариантны относительно преобразований Лоренца, не является уже ни математическим, ни даже дедуктивным. Это разрешение дилеммы, поставленной отрицательным результатом эксперимента Майкельсона — Морли (§ 16 гл. 1); из него следует, что все законы физики должны быть инвариантны относительно группы преобразований Лоренца.

Когда Эйнштейн в 1905 г. сформулировал эти новые для физики представления, идеи Феликса Клейна, касающиеся геометрии, были приняты и полностью оценены математиками того времени. Клейн изложил эти идеи в речи, прочитанной им при вступлении в должность профессора математики в Эрлангене. В этой речи, ставшей известной под названием Эрлангенской программы, он предложил рассматривать различные геометрии как изучение инвариантов соответствующих групп преобразований[1]). Выдающийся математик Герман Минковский, изумленный сходством между физическими идеями Эйнштейна и геометрическими идеями Клейна, сумел получить из них прекрасное сочетание — пространство-время, наделенное геометрией, в основу которой положены преобразования Лоренца.

Обсуждая роль математики в формулировании физических законов и выводе из них следствий, необходимо отметить часто возникающее несоответствие между глубиной физической теории и степенью сложности ее математического описания.

Математический аппарат специальной теории относительности предельно прост, в то время как лежащие в ее основе физические идеи и представления чрезвычайно тонки и глубоки. С другой стороны, многие проблемы, поставленные техникой, вносят незначительный вклад в наше понимание физического мира, однако требуют привлечения невероятно сложного математического аппарата. Кроме того, хотя (и это весьма примечательно) так часто какое-либо детище математики, задуманное и выращенное в ее недрах, оказывается неожиданно полезным для описания явлений внешнего мира (хорошими примерами служат комплексные числа и матрицы), тем не менее ни элегантность, ни особая сложность того или иного математического понятия, построения или метода сами по себе не дают никакой гарантии их практической полезности и пригодности.

---

[1]) См. Ф. Клейн, Сравнительное обозрение новейших геометрических исследований, сборник «Об основаниях геометрии», М., Гостехиздат, 1956, стр. 399—434. Речь Клейна была прочитана в 1872 г. — *Прим. ред.*

Вигнер так подытожил все это в своей статье «Непостижимая эффективность математики в естественных науках»[1]): «Математический язык удивительно хорошо приспособлен для формулировки физических законов. Это чудесный дар, который мы не понимаем и которого не заслуживаем. Нам остается лишь благодарить за него судьбу и надеяться, что и в будущих исследованиях мы сможем по-прежнему пользоваться им и что сфера его применимости (хорошо это или плохо) будет непрерывно возрастать, охватывая все более широкие области науки и принося нам не только радость, но и новые головоломные проблемы.»

Бесполезно было бы пытаться сколько-нибудь полно описать взаимодействие между математикой и физическими науками. Остановимся, однако, на одном важном аспекте этого взаимодействия, представляющем значительный интерес.

Внешний мир настолько сложен, что ученый-естествоиспытатель бывает доволен, если ему удается уловить и понять хотя бы некоторые самые простые из присущих миру закономерностей. Для этого он вводит упрощенные и идеализированные модели, освобожденные от маловажных и усложняющих дело подробностей и отражающие, как он надеется, наиболее существенные свойства рассматриваемых физических объектов.

Так, например, Ньютон при выводе законов Кеплера считал, что на планеты действует только притяжение Солнца. Он пренебрег действием других масс, хотя это, строго говоря, было неправильно. Позднее были предложены другие модели, более близкие к действительности. Одним из крупнейших достижений астрономии 19-го века было предсказание существования планеты Нептун, сделанное Адамсом и Леверье при попытке найти объяснение тому, что движение Урана заметно отклоняется от его кеплеровой орбиты.

Грубо говоря, дело обстоит так: вопрос о выборе модели решает ученый-естествоиспытатель; после этого

[1]) Русский перевод см. в книге Е. Вигнер, Этюды о симметрии, М., 1971, стр. 182—198.

выполняет свою роль математика, позволяющая дедуктивно выводить заключения уже только на основе предложенной модели. Все это достаточно хорошо известно и вряд ли требует дальнейшего обсуждения.

Существуют и модели иного типа, которые помогают разрешить логические трудности, возникающие при изучении других моделей, на вид вполне хорошо отражающих явления внешнего мира. Рассмотрим, например, тепловые явления при контакте двух тел A и B разной температуры, изолированных от всех остальных тел. Тогда, согласно законам термодинамики, должен возникнуть поток тепла только в одну сторону от более горячего тела (скажем, A) к более холодному (B) (*однонаправленный* поток).

В ходе этого процесса разность температур будет экспоненциально стремиться к нулю (закон теплопередачи Ньютона). Это следует из знаменитого второго начала термодинамики; одним из пессимистических следствий второго начала (в применении ко Вселенной) является полное выравнивание температур всех тел, которое Клаузиус назвал тепловой смертью.

Механический (кинетический) подход, при котором вещество рассматривается как совокупность частиц, а именно атомов или молекул, подчиняющихся обычным законам движения, приводит к совсем другой картине. Частицы, сталкиваясь друг с другом и двигаясь «случайным» образом, не могут создать *абсолютно однонаправленный* поток от A к B. Согласно теореме Пуанкаре, такая динамическая система в конце концов вернется в состояние, сколь угодно близкое к начальному, если только это начальное состояние не является столь исключительным, что такой возможностью можно спокойно пренебречь. Это «квазипериодическое» поведение резко отличается от монотонного стремления к выравниванию, которое следует из второго начала термодинамики.

Чтобы уладить возникшее расхождение, Пауль и Татьяна Эренфест предложили в 1907 г. простую и красивую модель (упомянутую в § 18 гл. 1).

Рассмотрим две урны A и B, одна из которых (скажем, A) содержит большое число N занумерованных шаров (в § 18 гл. 1 в качестве N было взято число $2R$). Сыграем теперь в такую игру: выберем «случайно» какое-нибудь число от 1 до N и переложим шар с этим номером из урны, где он лежит, во вторую (первым ходом всегда будет перекладывание из A в B). Затем повторим эту процедуру много раз (при этом шары будут часто возвращаться в A), следя за тем, чтобы последовательные вытягивания чисел были независимы и чтобы каждый раз извлечения всех чисел от 1 до N были равновероятными.

Интуитивно кажется, что до тех пор, пока в A намного больше шаров, чем в B, вероятность перекладывания из A в B будет значительно большей, т. е. получится нечто вроде однонаправленного потока из A в B.

Хотя вытягивания чисел и независимы, количества шаров в A в последовательные моменты времени не являются независимыми. Они связаны определенной зависимостью типа марковской цепи (см. § 18 гл. 1). *Среднее значение* числа шаров в A, экспоненциально убывая, стремится к $N/2$, что вполне согласуется с выводом термодинамики. С другой стороны, можно найти, что *с вероятностью* 1 *модель в конце концов вернется в начальное состояние* (т. е. все шары снова окажутся в A). Но в этом и состоит утверждение теоремы Пуанкаре для динамических систем.

Очевидно, что на самом деле между вторым началом термодинамики и квазипериодическим поведением динамических систем нет никакого противоречия, если только не рассматривать этот закон как абсолютную догму и допускать более гибкую интерпретацию, основанную на теории вероятностей. Все это становится еще яснее, если вычислить среднюю продолжительность интервала времени, необходимого для того, чтобы модель Эренфестов вернулась в начальное состояние. Для этого потребуется 2^N шагов — огромное число даже для не слишком больших N, скажем около 100.

И если все наблюдаемые явления кажутся нам необратимыми (однонаправленными), то только потому, что наша жизнь ничтожно коротка по сравнению с этими грандиозными сроками!

В «игру» Эренфестов легко играть при помощи современных вычислительных машин. Такие эксперименты проводились для $N = 2^{14} = 16\,384$ «шаров», причем каждый «прогон» состоял из 200 000 вытягиваний. (Это занимает меньше двух минут.) Число шаров в A регистрировалось после каждых 1000 вытягиваний. Один из полученных при этом графиков показан на рис. 34.

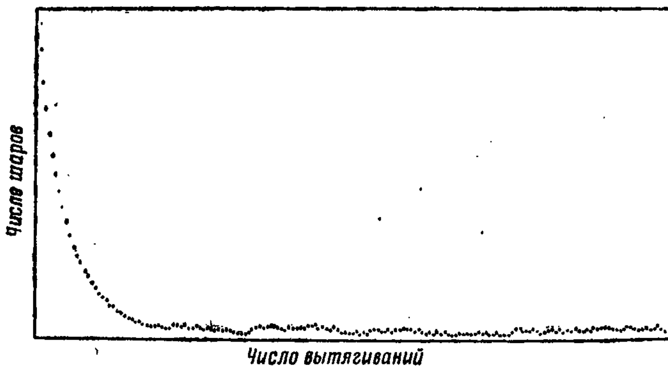


Рис. 34.

Как видно из этого графика, число шаров в A сначала падает почти в точности по экспоненте. Однако далее кривая становится «волнистой» и случайным образом колеблется относительно положения равновесия.

Как модель выравнивания температур модель Эренфестов весьма далека от реальности. И тем не менее именно она улавливает существо дела, позволяющее примирить кинетический подход с традиционной термодинамикой.

На протяжении 20-го века применение математических понятий, методов и технических приемов захватывает все больше областей знания и приложений. Можно даже отважиться на утверждение, что мы являемся свидетелями тенденции к «математизации» всех видов интеллектуальной деятельности. Такая тенденция, конечно, далеко не всегда оправдана. Можно назвать множество примеров, когда «математизация» тривиальна или претенциозна, и даже таких, когда она страдает обоими этими недостатками.

Однако, оставляя в стороне вкусы и личные точки зрения, невозможно отрицать, что число и разнообразие проблем, которые могут быть сформулированы и исследованы математически, постоянно увеличивается. Мы выделим из них и коротко обсудим здесь три проблемы, относящиеся соответственно к теории очередей, теории игр и теории информации.

Теория очередей возникла из попыток так спроектировать центральную телефонную станцию, чтобы каким-то образом свести к минимуму время ожидания связи. Опишем простейший тип возникающих при этом задач.

Допустим, что на станцию с одним обслуживающим аппаратом прибывают «клиенты» (поступают телефонные вызовы), которые обслуживаются (или обрабатываются) по очереди, один вслед за другим. Допустим также, что время можно разделить на элементарные интервалы продолжительности τ. («Квантовать» время здесь не обязательно, но если это сделать, задачу сформулировать легче. Решив ее в такой постановке, можно затем каким-то подходящим образом устремить τ к нулю и построить теорию, соответствующую случаю непрерывного прибытия клиентов.) Далее, обозначим через p_k вероятность того, что в течение некоторого данного интервала времени прибудет k $(k = 0, 1, 2, \ldots)$ клиентов. Тогда

$$p_0 + p_1 + p_2 + \ldots = 1.$$

Затем вводится существенное упрощение: предполагается, что прибытия клиентов в разные интервалы времени являются независимыми событиями, и, таким образом, вероятность того, что в течение первого интервала прибыло k_1 клиентов, в течение второго — k_2, третьего — k_3 и т. д., равна произведению

$$p_{k_1} p_{k_2} p_{k_3} \ldots .$$

Наконец, предполагается, что время обслуживания случайно, и вероятность того, что процесс обслуживания занимает время $\lambda\tau$ (т. е. λ элементарных интервалов, где $\lambda = 1, 2, \ldots$), обозначается через ρ_λ. Тогда

$$\rho_1 + \rho_2 + \rho_3 + \ldots = 1.$$

Теперь возникают следующие вопросы: каково среднее число клиентов, ожидающих своей очереди, по прошествии некоторого указанного времени? Каково среднее время, которое должен прождать клиент, прежде чем его обслужат? На эти вопросы получены полные ответы, которые, однако, отнюдь не являются простыми. Путь к ним неожиданно проходит по таким областям математики, как *теория функций комплексного переменного.* Например, приходится рассматривать степенные ряды

$$p_0 + p_1 z + p_2 z^2 + \ldots,$$
$$\rho_1 w + \rho_2 w^2 + \rho_3 w^3 + \ldots$$

для *комплексных* значений z и w.

Если рассматривать более близкую к действительности модель, допуская больше одного обслуживающего аппарата, математические трудности становятся почти непреодолимыми, и даже на простейшие вопросы невозможно ответить достаточно полно. К счастью, на помощь проектировщику сложной системы с несколькими обслуживающими аппаратами приходят быстродействующие вычислительные машины. Вдумчивое использование метода Монте-Карло (описанного в гл. 2) позволяет имитировать проектируемую систему и эмпирически исследовать различные стороны ее функционирования.

Строго говоря, такой «экспериментальный» подход не относится к математике. Однако эта ситуация похожа на ту, которая сложилась много веков назад, когда Евдокс и Архит пытались «подлить» немного механики в «светлые воды» геометрии (см. стр. 194—195).

Эмпирическое изучение очередей в сложных системах вполне может подсказать пути аналитического подхода, который потребует новых понятий и методов. Последние в свою очередь могут обогатить и украсить отдаленные и не связанные между собой области математики.

Теория игр, созданная Джоном фон Нейманом почти в одиночку, является удивительной иллюстрацией того, как можно «математизировать» задачи, которые на первый взгляд кажутся неподдающимися никакому рациональному подходу. Мы объясним, что это за теория, на примере упрощенного покера[1]).

Колода для игры в упрощенный покер состоит из $2n$ карт (n достаточно велико), половина которых — старшие (C), а вторая половина — младшие (M). Каждый из двух игроков A и B делает «ставку» размера a и получает одну карту. Затем A начинает игру. Он может либо «открыть» свою карту, либо «повысить» ставку, добавив в «банк» еще b денежных единиц. Если A открывает карту, то B обязан сделать то же самое. После этого игрок, у которого оказалась более сильная карта, забирает обе ставки; если же оба игрока имели одинаковые карты, то они делят банк между собой, т. е. каждый забирает назад свою ставку.

Если A «повышает» ставку, то у B уже есть выбор: он может либо «спасовать» (т. е. отказаться от игры и отдать деньги A), либо «играть» (т. е. добавить в банк ту же сумму b; в последнем случае A должен открыть свою карту, после чего карту открывает и B). Выигрыш, проигрыш и ничья определяются так же, как и в первом случае.

Вопрос заключается в том, какой способ игры наиболее выгоден для A (соответственно для B). ***Чистой стратегией*** называется правило, предписывающее, как должен поступить игрок в любой ситуации, которая может возникнуть в ходе игры. Таким образом, для A имеется четыре чистые стратегии:

(1) (О—О) — стратегия «открыть — открыть», т. е. открыть карту независимо от того, какая она у него: старшая (C) или младшая (M).

---

[1]) Этот пример, принадлежащий Таккеру, представляет собой упрощенный вариант примера, рассмотренного фон Нейманом и Моргенштерном. [См. также Дж. Кемени, Дж. Снелл, Дж. Томпсон, Введение в конечную математику, М., ИЛ, 1963, гл. VI, § 9. — *Ред.*]

(2) (О—П) — «открыть — повысить», т. е. открыть, если у него старшая карта, и повысить ставку, если его карта младшая.

(3) (П—О) — «повысить—открыть», т. е. повысить ставку, если он имеет старшую карту, и открыть карту, если она младшая.

(4) (П—П) — «повысить—повысить», т. е. повышать ставку в любом случае.

Следует заметить, что стратегии (О—О) и особенно (О—П) «плохие», ибо они не дают возможности использовать преимущество старшей карты.

Аналогично, B имеет четыре чистые стратегии, определяемые его решением «спасовать» (С) или «играть» (И) в разных ситуациях: (С—С), (С—И), (И—С), (И—И). (Напомним, что если A требует открыть карты, то у B нет никакого выбора.) Из этих четырех стратегий для B первая и вторая заведомо «плохие», так как они предписывают ему спасовать, имея на руках старшую карту.

Если предположить, что A и B играют ради выигрыша, а не ради скрытой благотворительности, то нужно сразу отбросить стратегии (О—О) и (О—П) для A, а также (С—С) и (С—И) для B. Теперь легко подсчитать, сколько может выиграть A при различных комбинациях стратегий. Допустим, например, что A выбирает стратегию (П—О) (т. е. повышает, если у него старшая карта, и открывает, если его карта младшая), а B — стратегию (И—И) (т. е. играет в любом случае). Тогда можно составить такую таблицу:

Карта A	Карта B	Выигрыш A
C	C	0
C	M	$a+b$
M	C	$-a$
M	M	0

При большом n четыре варианта исходных позиций — (C, C), (C, M), (M, C), (M, M) — будут осуществляться примерно с одинаковой частотой, равной $1/4$. Тогда «в среднем» стратегия (П—О) против (И—И) принесет A чистый выигрыш, равный $b/4$ за одну игру.

Аналогично можно подсчитать средний чистый выигрыш A за одну игру при выборе остальных трех пар стратегий. Результаты этих подсчетов можно изобразить в виде так называемой *матрицы платежа* (или матрицы выигрышей):

A \ B	(И – С)	(И – И)
(П – О)	0	$b/4$
(П – П)	$(a-b)/4$	0

Допустим, что $a < b$, т. е. левый нижний элемент этой матрицы отрицателен. Тогда стратегия (П—П) невыгодна игроку A, и он выберет стратегию (П—О). Аналогично, игроку B невыгодна стратегия (И—И), и он выберет чистую стратегию (И—С). Итак, в случае «консервативной» игры («повысить», если карта старшая; «открыть», если карта младшая; «играть», если карта старшая; «спасовать», если карта младшая) оба игрока в среднем будут «оставаться при своих». Оптимальными стратегиями являются чистые стратегии, и игра в этом случае честная (не дающая преимущества ни одному из игроков).

Если $a > b$, то результат игры смещается в пользу A, так как только он обладает привилегией «повышения»; в действительной игре это право предоставляется игрокам по очереди. Однако для того чтобы воспользоваться своим выгодным положением, A должен придерживаться *смешанной* стратегии, выбирая (П—О) с вероятностью p_1 и (П—П) с вероятностью p_2, где $p_1 + p_2 = 1$. Например, при $a = 8$ и $b = 4$

матрица платежа имеет вид

$$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

и A может *обеспечить* себе средний выигрыш в размере $^1/_2$ за одну игру, выбирая с вероятностью 50% стратегию (П—О) или (П—П). В свою очередь B, отвечая тем же (т. е. тоже выбирая свои стратегии с вероятностью 50%), может *помешать* A выиграть больше.

Отсюда видно, что в некоторых ситуациях для достижения оптимального результата игрок A в части играемых партий должен «блефовать» (т. е. повышать ставку, имея на руках младшую карту); в какой именно части партий ему следует это делать, видно из матрицы платежа.

Фон Нейман показал, что большой класс конфликтных ситуаций, подобных возникающим в экономике, можно рассматривать как *матричные игры*, т. е. игры, имеющие $(n \times m)$-матрицу платежа

$$\begin{pmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ \cdot & \cdot & \cdot & \cdot \\ a_{m1} & a_{m2} & \ldots & a_{mn} \end{pmatrix},$$

в которой элемент a_{ij} равен выигрышу игрока A, если он выбрал i-ю строку, а его противник B (втайне от A) выбрал j-й столбец.

Основная теорема теории игр утверждает, что существует такое число v, называемое *ценой* (или значением) игры, что A может *обеспечить* себе выигрыш, в среднем равный v за одну игру, в то время как B может *помешать* ему выиграть больше. Кроме того, существует оптимальная стратегия для A (вообще говоря, смешанная), гарантирующая ему выигрыш не меньше v за одну игру, и оптимальная стратегия для B (вообще говоря, тоже смешанная), гарантирующая, что его проигрыш за одну игру не превзойдет v.

По-видимому, еще рано судить о результатах применения теории игр, особенно в экономике, хотя именно там она нашла ряд наиболее известных (и наиболее разрекламированных) приложений. Одна из причин такой осторожности — огромные размеры матриц платежа в реальных ситуациях, так что полный их численный анализ все еще недоступен даже самым быстродействующим вычислительным машинам.

Важная роль теории игр определяется не только ее конкретными применениями в той или иной области знаний. Теория игр позволяет найти математический подход к целому ряду вопросов, связанных, если так можно выразиться, с рациональным поведением в конфликтных ситуациях. И даже если построенные на ее основе модели слишком упрощенны и нереалистичны, теория игр заслуживает большого доверия уже потому, что она дает надежду отыскать систематический подход к чрезвычайно сложным проблемам, связанным с общественным поведением.

Заканчивая обсуждение теории игр, невозможно не упомянуть хотя бы коротко теорию статистических решений, созданную на базе теории игр А. Вальдом.

Вальд рассматривал процесс принятия решения в условиях неопределенности как игру статистика против Природы. Стратегия Природы, конечно, неизвестна, однако статистик принимает решения в соответствии с оптимальной стратегией, которая определяется матрицей платежа. Эта матрица составляется из величин, которыми статистик оценивает для себя сравнительную стоимость того или иного решения. Эта теория по форме аналогична теории игр, однако технически гораздо более сложна и громоздка, так как матрицы платежа в большинстве случаев бесконечны. Влияние теории принятия решений на статистику было главным образом концептуальным. Теория статистических решений привлекла внимание ко многим важным вопросам, связанным со статистическими выводами, и особенно с характером статистических критериев, и внесла в них известную ясность.

Теория информации занимается проблемами, связанными с эффективностью передачи сообщений. Типичная ситуация здесь состоит в том, что имеется источник информации, выбирающий из некоторого множества сообщений *одно сообщение*, которое должно быть передано; передающее устройство превращает это сообщение в *сигнал*; далее, имеется канал (линия связи), по которому посылается сигнал, и, наконец, принимающее устройство, преобразующее сигнал в сообщение. Например, при передаче телеграмм записанные буквами слова кодируются последовательностями импульсов тока переменной длительности (тире, точки, пробелы), которые передаются по проводам и затем снова преобразуются в составленные из букв слова.

Теория информации не занимается проблемами семантики (насколько полно передаваемые символы отражают смысл сообщения); в ней рассматриваются только вопросы, связанные с безошибочностью (точностью) и экономичностью передачи.

Чтобы пояснить, какого рода задачи возникают в теории информации, допустим, что сообщение представляет собой строку из N букв латинского алфавита (N достаточно велико), в которой каждая буква встречается с той же частотой, с какой она появляется в «среднестатистических» текстах на английском языке. Можно представить себе и более общую ситуацию, когда имеется алфавит из k букв $S_1, S_2, \ldots, S_k$, причем появление в сообщении буквы S_1 имеет вероятность p_1, буквы S_2 — вероятность p_2 и т. д. Следующие одна за другой буквы выбираются независимым образом. Такие сообщения можно передавать последовательно, буква за буквой. Допустим, что передача одной буквы занимает одну единицу времени (скажем, одну микросекунду); тогда скорость передачи равна одному символу за единицу времени. Нельзя ли улучшить положение? Какова максимальная скорость, с которой может быть осуществлена передача?

Передача по буквам неэффективна: при такой передаче не используется то важное обстоятельство, что некоторые сообщения выбираются источником значительно реже, чем другие. Скорость передачи можно повысить, закодировав частые сообщения короткими выражениями и оставив более длинные выражения более редким сообщениям.

Шеннон ввел в рассмотрение две величины: H — *энтропию* источника и C — *пропускную способность* канала и доказал, что оптимальная скорость передачи равна отношению C/H, которое всегда не меньше единицы. Это означает, что можно придумать коды, позволяющие осуществлять передачу с любой средней скоростью, меньшей, чем C/H, и не существует кодов, обеспечивающих большую, чем C/H, скорость передачи сообщений.

Энтропия H определяется (грубо говоря) как

$$-\frac{1}{N}\cdot(\text{логарифм вероятности «типичного» сообщения}).$$

Пропускная способность C равна максимуму H по всем возможным заданиям вероятностей, совместимым с ограничениями, которые налагаются на сообщения. В рассматриваемом простом случае, когда сообщение представляет собой строку из N букв, выбираемых независимым образом, на источник не налагается никаких ограничений. К вопросу об ограничениях мы вернемся несколько ниже.

Для теоретических целей достаточно рассматривать только *двоичные коды* (т. е. коды, представляющие собой последовательности нулей и единиц). Поэтому в теории информации удобно пользоваться логарифмами при основании 2. Это, конечно, не вызвано необходимостью: подобное соглашение равнозначно, скажем, выбору системы единиц.

Чтобы получить представление о том, что понимается под «типичным» сообщением, вернемся к нашему примеру.

Если N велико, то бо́льшая часть сообщений содержит *приблизительно* p_1N букв S_1, p_2N букв S_2 и т. д. Это утверждение является грубой формулировкой закона больших чисел, рассмотренного в главе 1. Типичным считается сообщение, которое *действительно* содержит p_1N букв S_1, p_2N букв S_2 и т. д. Числа p_1N, p_2N, ..., конечно, могут быть и не целыми; в таком случае нужно брать ближайшие к ним целые числа; в пределе при $N\to\infty$ замена, скажем, p_1N целым числом $[p_1N]$ не отразится существенно на результате.

Вероятность того, что сообщение из N букв будет содержать p_1N букв S_1, p_2N букв S_2 и т. д., равна

$$p_1^{p_1N} p_2^{p_2N} \ldots p_k^{p_kN},$$

и, следовательно, в нашем простом примере энтропия задается формулой

$$H = -(p_1 \log p_1 + p_2 \log p_2 + \ldots + p_k \log p_k),$$

где $p_1, p_2, \ldots, p_k$ удовлетворяют единственному условию

$$p_1 + p_2 + \ldots + p_k = 1.$$

Максимальное значение H при таком условии равно

$$H_{\max} = C = -\log k;$$

оно получается, когда все p_i равны между собой. Следовательно, можно построить код, обеспечивающий любую скорость передачи, меньшую чем

$$\frac{\log k}{p_1 \log p_1 + \ldots + p_k \log p_k}.$$

До сих пор мы задавали только частоту появления в сообщении каждого символа в отдельности. Однако в конкретном языке, скажем английском или французском, на последовательности букв, образующие допустимое сообщение, налагаются очень жесткие ограничения, часть которых не известна. Реальный язык можно аппроксимировать, налагая все больше и больше ограничений статистического характера на процесс генерации сообщений. Например, вместо условия независимости можно ввести требование, чтобы каждая диграмма (т. е. каждая *пара* из двух последовательных букв) появлялась в сообщениях с той же частотой, что и в реальных текстах на данном языке; тем самым будет достигнуто большее соответствие с действительной структурой языка. Этот процесс можно продолжить, подгоняя частоты троек, четверок и т. д. последовательных букв к реальным частотам соответствующих сочетаний в данном языке. Если ограничиться диграммами, то статистическое описание источника сообщений примет вид простой марковской цепи.

В применении к нашему искусственному примеру с алфавитом $S_1, S_2, \ldots, S_k$ это означает, что заданы вероятности p_{ij} того, что за S_i следует S_j, и вероятность сообщения $S_{i_1}S_{i_2}\ldots S_{i_N}$ равна

$$p_{i_1i_2}p_{i_2i_3}\ldots p_{i_{N-1}i_N}.$$

Вероятности $p_1, p_2, \ldots, p_k$ появления отдельных символов в длинных сообщениях можно найти, решая линейные уравнения

$$\sum_{i=1}^{k} p_i p_{ij} = p_j \quad (j = 1, 2, \ldots, k).$$

Энтропия такого источника равна

$$H = -\sum_i p_i \sum_j p_{ij} \log p_{ij}.$$

Можно показать, что эта величина не превосходит энтропии источника в случае независимой генерации символов с вероятностями p_i. В этом проявляется общий принцип: чем больше налагается ограничений, тем меньше становится энтропия.

Если некоторое p_{ij} равно 0 или 1 (например, в английском языке за буквой *z никогда* не следует *x*, так что $p_{zx} = 0$), то мы имеем *абсолютное* ограничение. При отыскании максимума энтропии можно варьировать все вероятности p_{ij}, кроме тех, которые равны 0 или 1.

До сих пор мы предполагали, что в канале отсутствует шум, т. е. что каждый символ передается абсолютно точно. Наиболее интересные математические задачи возникают в ситуациях, когда канал «зашумлен». Простейшая модель такого зашумленного канала — двоичный канал без памяти. Здесь мы считаем, что при передаче двоичных кодов имеется некоторая постоянная вероятность p того, что символ 0 или 1 будет передан правильно, и постоянная вероятность $q = 1 - p$ того, что он будет искажен (т. е. 0

заменится на 1 или 1 на 0); кроме того, мы полагаем, что отдельные символы передаются независимо.

Шеннон и другие показали, как и при каких обстоятельствах можно построить коды, допускающие дешифровку с произвольно высокой вероятностью; найдены также оптимальные скорости передачи.

Эти разделы теории уже чрезвычайно сложны, но даже из нашего краткого и неполного обзора ее более элементарных частей видно, с каким успехом математика применяется сейчас к задачам, которые совсем недавно считались недоступными никакому точному количественному анализу.

Обсуждая связи математики с другими науками, нельзя не коснуться статистики. Статистика не является ветвью математики, поскольку она занимается обработкой данных и принятием решений на основе результатов этой обработки. Используемая таким образом, она *не является* даже четко очерченной дисциплиной, а скорее представляет собой общий инструмент научного исследования. Однако математика играла и играет важную роль в развитии статистики. Многие разделы статистики настолько глубоко пропитаны математическими идеями и методами, что их совокупность получила наименование математической статистики.

В свою очередь статистическая точка зрения оказывается полезной во многих областях чистой математики, расширяя проблематику и подсказывая новые пути и подходы.

Мы хотим снова подчеркнуть, что очень редко можно провести четкую границу между математикой и другими науками, к которым она применяется. Попытки — к сожалению, довольно частые — изолировать «чистую» математику от всей остальной научной деятельности и заставить ее вариться в собственном соку могут лишь обеднить и математику, и прочие науки.

ГЛАВА 4

ИТОГИ И ПЕРСПЕКТИВЫ

Маршруты наших путешествий в математику мы выбирали исходя из ее истории, внутренних связей ее разделов и развития в ней синтезирующего начала. Мы рассмотрели задачи о целых числах, в связи с которыми родилась идея бесконечности, а потом перешли к примерам из геометрии, проследив предварительно эволюцию абстрактных представлений о числах и геометрических объектах. Мы попытались, по возможности просто, показать, каким путем математики пришли к рассмотрению общих групп преобразований, а затем, исследуя множества таких объектов, как пространства, — к построению теорий общих структур. Хотя разнообразие математических объектов в наше время огромно, сам математический метод остался таким же, каким был всегда: сначала постулируется или молча принимается небольшое число аксиом, а затем путем повторного применения определенных правил (математической логики) строится теория, т. е. совокупность теорем, описывающих свойства и отношения между объектами, удовлетворяющими этим аксиомам. Воскресни сегодня такие математики, как Архимед, Евклид или Ньютон, они могли бы растеряться от обилия понятий, интересующих современных математиков, однако наши методы они нашли бы вполне понятными и знакомыми.

В столь кратком очерке мы поневоле должны были ограничиться отдельными избранными темами и методами современной теоретической и прикладной математики; многие из новейших методов нам не удалось даже упомянуть. В этой заключительной главе мы хотим назвать ряд областей математики, где ведутся особенно интенсивные исследования, и обрисовать процесс растущей математизации различных отраслей науки и техники.

Все самые выдающиеся достижения математического метода связаны с тем, что он позволяет абстрагировать определенные свойства наблюдаемых объектов и наблюдаемых отношений между ними и чисто логическим путем выводить новые свойства и новые отношения, которые можно затем проверить наблюдением и экспериментом. Так, сформулированные Ньютоном законы механики позволили чисто математическими средствами возвести грандиозное здание классической механики и найти законы движения небесных тел. Тем же духом проникнуты и все дальнейшие успехи математической физики, открывшие дорогу для развития других областей науки и техники. В самой математике новая теория обычно начинается с того, что постулируется ряд новых математических свойств. Так возникала теория вероятностей, различные геометрии, теория аналитических функций, теория пространств, элементами, или «точками», которых служат функции. Этот процесс аксиоматического построения новых теорий продолжается и поныне. Наблюдая отдельные классы явлений, мы выделяем из них путем абстракции более простые «фундаментальные» классы, постулируем те или иные свойства этих классов и выводим математические следствия из построенных таким образом моделей. Одновременно изучаются и сравниваются свойства различных классов и прилагаются усилия к тому, чтобы свести их воедино в рамках новых создаваемых для этого «сверхтеорий». Иначе говоря, в процессе совместной деятельности математиков, из их общих интересов и результатов рождаются новые математические концепции, которые в дальнейшем сами оказываются частными случаями более общих закономерностей.

Мы попробуем описать современные исследования на примере нескольких таких новых теорий. Элегантную и логически связную часть математики составляет созданная Шенноном и его последователями *теория информации*. Выше, когда мы говорили о теории информации, речь шла о конечном множестве событий и приписанных им вероятностях. Интересно, что понятия теории информации можно определить и для бесконечного пространства событий (как дискретного, так и непрерывного); для этого нужно ввести меру в таком пространстве событий либо путем предельных переходов, либо путем интегрирования (вспомним сказанное выше в параграфе, посвященном теории меры). Кроме меры, удалось ввести и другие характеристики множеств в пространстве событий. Если в пространстве событий определено расстояние между элементами, то можно определить энтропию, или емкость, множеств в этом пространстве. Эти определения, выросшие из практических задач передачи и кодирования сообщений, позволили математикам развить общие абстрактные теории, которые помогли решить несколько давних проблем теории функций. В частности, больших успехов в применении идей Шеннона для решения задач чистой математики добились советские ученые. Удивительно, как много времени ушло на то, чтобы обобщить идею энтропии, перейдя от первоначального понятия, относившегося к совокупностям молекул или атомов, к формулировке, охватывающей весьма общие классы событий. Помимо той пользы, которую это обобщение принесло для решения проблем теории связи, оно оказалось исключительно ценным для решения ряда абстрактных математических проблем, на первый взгляд не имевших, как будто бы, ни малейшего отношения к идеям теории вероятностей.

Выше мы уже упоминали знаменитый список проблем Гильберта. Одна из этих проблем — найти выражение корней алгебраического уравнения степени n в виде функций от его коэффициентов. Известно, что каждому такому уравнению соответствует однозначно определенное множество корней; следовательно, эти корни должны быть функциями его коэффициентов. Для уравнений первой, второй, третьей и четвертой степени эти функции имеют весьма специальный вид: они получаются из коэффициентов путем сложения, умножения и извлечения корней. Функции такого вида от n переменных $(n = 1, 2, 3, 4)$ можно представить в виде суперпозиции функций меньшего числа (двух) переменных. В самом деле, сумму или произведение любого числа членов можно найти, выполняя нужное число раз сложение или умножение двух членов, а операция извлечения корня степени k определяет функцию одной переменной. При $n \geqslant 5$, согласно знаменитому результату Галуа, корни алгебраического уравнения не выражаются через его коэффициенты в виде радикалов и рациональных функций. Проблема, поставленная Гильбертом, заключается в следующем: если сложения, умножения и извлечения корней недостаточно для того, чтобы выразить корни алгебраического уравнения степени $n \geqslant 5$ в виде функций коэффициентов, то нет ли каких-нибудь других функций меньшего числа переменных, путем повторной суперпозиции которых можно получить эти корни, т. е. решить уравнение? Более общая формулировка этой проблемы Гильберта такова: всегда ли возможно выразить непрерывную функцию многих переменных в виде суперпозиции функций меньшего числа переменных, скажем всего двух? Хотя сам Гильберт был склонен считать это невозможным, А. Н. Колмогорову и В. И. Арнольду удалось доказать, что всякую непрерывную функцию любого конечного числа действительных переменных можно представить в виде суперпозиции непрерывных функций не более двух переменных. Этот результат, полученный в 1957 г., позволил уточнить и обобщить формулировки других проблем, связанных с суперпозициями функций. Речь идет о проблемах аналогичного представления *наборов n* функций от n переменных: можно ли выразить взаимно однозначное непрерывное преобразование n-мерного пространства в виде суперпозиции непрерывных преобразований с меньшим числом переменных? При каких условиях это возможно? Вместо одной только непрерывности можно наложить более сильные ограничения, потребовав, чтобы функции или преобразования были дифференцируемыми, аналитическими и т. д.

Выше мы видели, как удобно на языке преобразований формулировать качественные свойства движений физических систем. Динамическая система из n материальных точек представляется одной точкой $6n$-мерного пространства, а изменения этой динамической системы во времени описываются движениями этой точки. Совокупность всех возможных начальных положений, изменяющихся во времени, определяет поток в $6n$-мерном фазовом пространстве. «В общем случае» такой поток, если он сохраняет объем или меру, является эргодическим, т. е. фазовая точка с одинаковой вероятностью может попасть в любую часть доступного пространства. Поясним, что означают здесь слова «в общем случае». Множество всех непрерывных сохраняющих меру преобразований можно рассматривать как пространство, элементами, или «точками», которого служат эти преобразования. В этом пространстве можно определить расстояние между точками. Множество эргодических преобразований в этом пространстве является «большим» в том смысле, что его дополнение — множество преобразований, не являющихся эргодическими, «мало»: его можно представить в виде объединения счетного числа множеств, нигде не плотных во всем пространстве. Здесь оказывается полезным понятие функционального пространства; хотя это понятие было введено в результате чисто абстрактных рассуждений, оно во многих случаях помогает формулировать точные утверждения о физических системах. Вывод о том, что среди всех сохраняющих объем непрерывных потоков «большинство» эргодичны, напоминает утверждение, что «большинство» действительных чисел иррациональны или даже трансцендентны. Конечно, то или иное число, определенное какими-то уравнениями или алгоритмом, не обязательно должно принадлежать этому «большому» множеству. В каждом конкретном случае нелегко установить, обладает ли данная динамическая система свойством эргодичности, кроме тех случаев, когда движение динамической системы определяет непрерывный сохраняющий объем поток весьма *специального* вида. Дж. Мозер в США и А. Н. Колмогоров и В. И. Арнольд в СССР выявили классы динамических систем, фазовые точки которых движутся не по всему пространству, а описывают квазипериодические траектории, не выходящие за пределы некоторых характерных участков пространства. Иными словами, эти физические системы обладают свойствами, представляющими собой нечто промежуточное между свойствами простого периодического движения (типа кеплеровой системы двух тел) и свойствами «общего» непрерывного потока. Раньше предполагалось, что если рассматривается «достаточно усложненная» система, то ее движение стремится стать эргодическим, т. е., грубо говоря, по прошествии достаточного времени система будет близка к любому возможному положению. Однако оказалось, что некоторые специальные системы могут и не обладать этим свойством. Например, если движение можно описать *линейными* уравнениями, как в случае механических колебаний какой-нибудь физической системы, то (по крайней мере при малых амплитудах) мы получим периодические колебания. Так, движение идеально упругой струны всегда будет складываться из периодических колебаний, соответствующих ее собственным частотам. Разумеется, линейные уравнения являются лишь приближением реальной физической ситуации. Эмпирически сила упругости не будет строго пропорциональна смещению. Выражение этой силы как функции смещения включает, кроме главного члена, которому эта сила пропорциональна, еще ряд малых членов — второго или более высоких порядков. Принимая это во внимание, можно было бы думать, что с течением времени первоначальная форма колебания будет постепенно становиться все более и более сложной. Расчеты на вычислительной машине позволили имитировать это движение за достаточно продолжительный промежуток времени. Выяснилось, что вопреки всем ожиданиям колебания не становятся чрезвычайно сложными и что струна колеблется хотя и не вполне периодически, но в пределах небольшой части допустимого класса положений. Иными словами, эта система не обладает свойством эргодичности. Результаты этих расчетов послужили толчком к широкому исследованию такого рода «нелинейных» задач, в том числе и таких, которые представляют чисто математический интерес. Как мы видели выше, о линейных преобразованиях евклидова пространства известно уже очень многое; преобразования же такого пространства в себя, не являющиеся линейными, изучены мало. Можно ожидать, что из исследований в этом направлении будут постепенно вырисовываться контуры новой общей теории.

Теория игр (с которой мы уже немного познакомились раньше) связана с особого рода математическими задачами из области комбинаторики. Представим себе двух игроков, по очереди выбирающих ходы из какого-то заданного множества альтернатив. После некоторого числа ходов возникает ситуация, которая рассматривается как «выигрыш» одного из партнеров. В теории игр изучается выбор стратегий при условиях, когда каждый из игроков должен в своих решениях основываться на вероятностях различных решений противника. Прежде всего здесь возникает задача оптимизации тактики каждого из партнеров. Большая часть теории посвящена играм «с неполной информацией», в которых значительная роль приходится на долю случая. Так обстоит дело, скажем, в покере в отличие от шахмат, являющихся типичным примером игры «с полной информацией». Изучаются также игры, в которых участвует более двух партнеров; важной проблемой здесь является образование *коалиций* одних игроков против других. Начало этим исследованиям в их современной форме было положено статьей Дж. фон Неймана, а свое дальнейшее развитие они получили в книге фон Неймана и Моргенштерна[1]), давшей толчок целой серии последующих работ.

Классическим примером «ясновидения», которым может обладать математическое воображение, предвосхищая развитие физических теорий, служит

---

[1]) Дж. фон Нейман, О. Моргенштерн, Теория игр и экономическое поведение, М., «Наука», 1970. Первое английское издание книги вышло в 1944 г. — *Прим. ред.*

создание римановых геометрий. Б. Риман в своей знаменитой лекции «О гипотезах, лежащих в основании геометрии» (Über die Hypothesen die der Geometrie zugrunde liegen)[1]) определил целый класс геометрий, рассматривая в качестве обобщения «плоского» евклидова пространства «искривленные» многообразия, где кривизна задается локально, в окрестности каждой точки пространства; например, она может определяться какой-нибудь физической величиной типа плотности вещества. Пророческий характер этой новой теории обнаружился много лет спустя, когда Эйнштейн положил ее в основу общей теории относительности. Аппарат дифференциальной геометрии, разработанный Риманом и другими математиками после него, был существенным образом использован для формулировки принципов общей теории относительности. Обе эти теории носят локальный характер, интересуясь прежде всего поведением кривизны и геодезических (т. е. кратчайших линий, соединяющих точки) в окрестности каждой точки. Глобальные свойства таких пространств связаны с топологическими характеристиками пространства в целом. Такими характеристиками являются, например, число k-мерных дырок в пространстве (числа Бетти) и группы гомотопий пространства (порождаемые не стягиваемыми друг в друга кривыми и поверхностями в этом пространстве). Многие важные исследования посвящены изучению подобных топологических свойств пространств; дифференциальная геометрия «в целом» составляет один из самых активно разрабатываемых разделов современной математики. Применяемые для этого методы, алгебраические по своему характеру, позволяют изучать свойства непрерывных векторных полей на таких пространствах.

В духе того же глобального подхода математики интенсивно изучают структуру непрерывных групп.

---

[1]) См. Б. Риман, Сочинения, М.—Л., Гостехиздат, 1948, стр. 279—293, 509—526, или сборник «Об основаниях геометрии», М., Гостехиздат, 1956, стр. 308—341. Лекция Римана была прочитана в 1854 г. — *Прим. ред.*

Например, группы вращений n-мерных пространств сами можно рассматривать как пространства, определив в них расстояние между любыми двумя вращениями. В последнее время удалось выяснить многие топологические характеристики таких групп, особенно в случае, когда групповая операция не только непрерывна, но и дифференцируема. Такие группы называются группами Ли. Для них ищут представления группами линейных преобразований n-мерного пространства, т. е. ищут группы линейных преобразований, изоморфные группам Ли. Математические свойства таких представлений получили важную физическую интерпретацию в квантовой механике и теории элементарных частиц. Применение этих идей к классификации атомных спектров и элементарных частиц — еще одно свидетельство поразительных возможностей математического предвидения.

Непрерывный рост количества отдельных и частных результатов, многочисленные, но несогласованные попытки унифицировать математику и, наконец, растущая математизация науки и техники ставят перед математикой серьезную проблему возможной потери общего языка: обнаруживается явная тенденция к вавилонскому столпотворению.

В этой связи полезно в историческом плане проследить, какую роль сыграла математика в развитии других наук. Не случайно, что дифференцирование и интегрирование были придуманы тогда же, когда Ньютон открыл законы механики и обосновал законы движения небесных тел. В то самое время, когда были сформулированы все основные законы механики, были изобретены и усовершенствованы и средства для вывода следствий из этих законов, причем до сих пор, по-видимому, не существует других, теоретически более совершенных или технически более эффективных средств для формулировки этих законов и для расчетов движения тел. Дифференциальные и интегральные операторы и сегодня составляют основу математического анализа. Законы классической физики формулируются в виде дифференциальных уравнений или систем таких уравнений. Сначала это были *обыкновенные* дифференциальные уравнения: они связывают производные неизвестной функции одной переменной со значениями самой этой функции и других заданных функций. При таком подходе поведение физической системы описывается в терминах поведения материальных точек, которые моделируют фактическое распределение масс.

Для математического описания непрерывных распределений масс или полей требуются дифференциальные уравнения *с частными производными*. В этом случае рассматриваются функции нескольких переменных, а в уравнения входят частные производные этих функций по пространственным координатам и по времени. Подобные уравнения применялись последователями Ньютона еще в 18-м веке. Примерами описываемых такими уравнениями функций могут служить скорости в жидкостях, плотность вещества в пространстве, упругие напряжения в материале, зависимость температуры от координат и времени.

При помощи дифференциальных уравнений с частными производными были поставлены и решены многие задачи гидродинамики, теории упругости и теории теплоты. На протяжении 19-го столетия математическая физика добилась ряда крупных достижений путем применения математического анализа. Несколько позднее работы по теории электричества увенчались математическим описанием электромагнитных явлений, полученным Максвеллом, выразившим основные законы электромагнетизма в виде системы дифференциальных уравнений с частными производными.

В прошлом веке использование в физике функций комплексной переменной прямо-таки чудодейственным образом помогло создать эффективные алгоритмы решения задач, которые до этого не удавалось решить никакими другими методами. Мало того, оказалось, что благодаря этому физические законы получили новый смысл и новые формулировки, что уж совсем похоже на мистику: ведь, как мы помним, комплексные переменные (и функции с комплексными значениями) первоначально возникли в алгебре, вне всякой связи с проблемами естественных наук.

Еще одна область математического анализа, начало которой было положено в 18-м веке, — это *вариационное исчисление.* Физические законы можно формулировать как утверждения о том, что некоторые интегралы функций одной или нескольких переменных принимают *экстремальное* значение. Компактны, изящны и обладают большой математической силой принцип Ферма (о кратчайшем времени прохождения светового луча через оптическую систему) или принцип наименьшего действия (в форме Гамильтона или Мопертюи — Лагранжа). В конце прошлого столетия возникла теория *интегральных уравнений,* отдельные примеры которых изучал еще Абель. Все эти исследования относятся к математическому анализу.

Свою дорогу в физику нашли и другие разделы математики. Теория вероятностей образовала фундамент *статистической механики,* которая изучает поведение вещества на основе дискретной, а не непрерывной, математической модели, рассматривая его как совокупность огромного количества взаимодействующих частиц. Такой «комбинаторный» подход существует параллельно с непрерывным подходом термодинамики и дополняет его. Наконец, как мы уже говорили, развитие в прошлом веке теории новых геометрий подготовило создание специальной и общей теории относительности.

Эти новые применения математики в физических теориях, включая великое творение Эйнштейна — теорию относительности, были найдены уже в 20-м веке. Весь необходимый математический аппарат теории относительности был подготовлен еще до Эйнштейна. Лоренц и Пуанкаре исследовали группу преобразований четырехмерного пространства-времени, относительно которых инвариантны уравнения Максвелла, описывающие электромагнитные явления. Инвариантность относительно этих преобразований всех уравнений, описывающих физические явления, Эйнштейн возвел в основополагающий физический принцип; это означало радикальный переворот в существовавших тогда представлениях о пространстве и времени. Такие ошеломляющие выводы, как утверждение об эквивалентности массы и энергии, были получены как *математические следствия* этого допущения. Формула $E = mc^2$ является математическим следствием инвариантности законов природы относительно преобразований Лоренца. Проделанная математиками работа по определению абстрактных понятий и развитию абстрактных идей существенно облегчила разработку новых математических схем для теоретической физики, а в ряде случаев непосредственно способствовала их созданию. Так, например, работа Римана и других математиков над геометриями, более общими, чем евклидова, подготовила почву и создала математический аппарат общей теории относительности.

В квантовой механике различные явления рассматриваются с точки зрения, которая представляется еще более абстрактной. Основными объектами — «исходными понятиями» этой теории служат уже не материальные точки евклидова пространства, а функции распределения, характеризующие «волновые пакеты». Они рассматриваются как первооснова физического объекта, причем наблюдаемыми характеристиками физических явлений оказываются интегралы этих распределений или производные от них. Математической основой этой теории служит теория функциональных пространств, подобных гильбертову пространству, упомянутому в главе 2. Задолго до создания квантовой механики Гильберт, определяя математические свойства линейных преобразований своего бесконечномерного пространства, употребил слово «спектр». Оказалось, что этот математический спектр в точности соответствует спектру излучения атома!

Кажущийся хаос спектральных линий можно понять и упорядочить с помощью математической теории групп. В самой математике идея о том, что формальные свойства групп преобразований могут служить для определения и классификации объектов, на которые действуют эти преобразования, была впервые применена в геометрии. Ф. Клейн в своей знаменитой Эрлангенской программе провозгласил, что любая геометрия определяется группой преобразований, относительно которых инвариантны ее объекты и отношения между ними. В настоящее время наблюдаются тенденции к обобщению этого подхода, и многое из того, что сейчас делается в теоретической физике, можно считать развитием этой идеи. Группы преобразований использовались для вывода законов сохранения и классификации элементарных частиц. С помощью абстрактных групп изучаются законы сохранения импульса и энергии, а также сохранения заряда и таких величин, как «спин» и «странность».

Какие из математических теорий, вероятнее всего, будут играть важную роль в дальнейшем развитии физических теорий? Явления чрезвычайно малых масштабов, происходящие на внутриатомном и ядерном уровне, совершенно не согласуются с представлениями классической физики. Даже для чисто качественного их описания требуются математические переменные иного типа, чем привычные действительные числа и евклидовы континуумы. Оказывается невозможным с произвольной точностью одновременно измерить импульс *и* координаты частицы или энергию *и* время испускания излучения. По мере того как, изучая строение вещества, мы будем переходить ко все более высоким энергиям, могут потребоваться математические модели, совсем непохожие на те, которые применяет современная физика. Полезной для построения модели вещества и излучения может оказаться теория множеств (и топология точечных множеств). Перемены будут еще более радикальными, если подобные идеи будут использованы для построения моделей самого пространства-времени. Последние работы в области астрономии и космологии — науки, рассматривающей Вселенную в целом, обнаружили ряд неожиданных явлений. Если выяснится, что большей адекватностью обладают модели реально бесконечной Вселенной, то это повысит роль математических идей теории множеств. Последние открытия математической логики относительно неполноты любой системы аксиом ставят как строго научные, так и философские проблемы о природе Вселенной. Мы уже рассказывали о том, как в последние годы в самой математике было доказано, что некоторые важные проблемы неразрешимы в существующих аксиоматических системах. Может статься, что никакая конечная система аксиом никогда не будет признана определенной или окончательной. Подобные дилеммы вырастают не только из чисто умозрительных построений (в частности, математических), но и из того, что мы называем физическим миром. Если Вселенная действительно содержит бесконечное множество дискретных «точек» (будь то звезды, элементарные частицы или фотоны), то должны существовать высказывания или предложения об этих множествах, неразрешимые в терминах любого конечного числа заранее сформулированных законов и правил. Вот какие серьезные последствия, быть может, вытекают из этих недавних математических работ.

Мы уже много говорили об органической связи между математикой и математической физикой. Ни одна из этих двух важнейших сфер человеческой мысли не была бы даже отдаленно похожа на то, чем она стала сегодня, если бы она постоянно не подвергалась влиянию другой. Самая суть теоретической физики заключена в ее математических формулировках; с другой стороны, и развитие ряда важных разделов математики стимулировалось и определялось проблемами, вытекавшими из наблюдений за поведением вещества и излучения. Наши представлеиия о пространстве и времени являются абстракциями, основанными на эмпирическом опыте. Каким образом получается, что этот опыт поддается математической интерпретации, которую можно затем логически развивать и в итоге приходить к выводам, согласующимся с данными наблюдений, — это, быть может, сложный философский вопрос. Одной из причин, почему это возможно, является обязательное требование, чтобы все измереиия, а тем самым и большую часть результатов физических и астрономических исследований, можно было свести к операциям над числами.

Совершенно по-иному обстоит сейчас дело с науками, изучающими живую материю, которая характеризуется гораздо большим богатством и многообразием форм. Происходящий в иаши дии бурный рост знаний о первичных, или «элементарных», биологических явлениях делает эти явления все более доступными для исследования математическими методами. В биологии такие методы уже давно и с большим успехом используются для решения многих частных технических задач. Для решения задач, связанных со статистическим поведением (например, при изучении химических реакций в живом организме), с закономерностями поведения больших совокупностей органических компонентов и их организации, успешно применяются дифференциальное и интегральное исчисление, алгебра и комбинаторика. Такие работы, как исследование Вольтерры об изменении численности особей в биологических видах, из которых одни питаются другими, помимо своего значения для биологии, представляют и математический интерес. Вольтерра использовал систему нелинейных обыкновенных дифференциальных уравнений. Его работа послужила толчком для изучения нелинейных проблем в чистой математике; за последнее время в этой области получены многие важные результаты. В связи с законами наследственности, вскрытыми Менделем, был проведен ряд комбинаторных исследований. Математические методы существенным образом используются для описания поведения различных смесей в биохимии, а для изучения термодинамических и квантовых основ подобных процессов не только полезен, но просто необходим весь сложный аппарат математической физики.

Более того, мы уверены, что достижения биологии за последний десяток лет открывают еще более широкие, захватывающие и многообещающие математические перспективы.

В последние годы получены первые важные результаты, касающиеся механизмов функционирования живой клетки. Общепринятая геометрическая модель ДНК, предложенная Криком и Уотсоном, представляет собой длинную спиральную двойную структуру, построенную из четырех видов оснований, с поперечными связями между цепями. Предполагается, что в процессе воспроизведения эта «лесенка» (двойная спираль) расщепляется на две цепи. Составные элементы каждой из этих цепей находят для себя комплементарные элементы в окружающей среде; таким образом из двух одинарных цепей ДНК получаются две новые двойные цепи, идентичные исходной двойной спирали. Последовательность четырех оснований, из которых состоит молекула ДНК, служит генетическим кодом клетки и всего организма. В этой последовательности закодирована, в частности, информация о биосинтезе белков. Кроме того, в ней, по-видимому, содержатся и инструкции (с логической точки зрения более высокого порядка) о *функциональном* поведении, нечто вроде общей «блок-схемы» для вычислительной машины. Другие находящиеся в клетке молекулы, по-видимому, получают все эти инструкции от ДНК и передают их далее, туда, где происходят процессы биосинтеза.

В понимании механики, логики и комбинаторики этих процессов полной ясности достичь пока что не удалось. Несомненно во всяком случае, что при математическом анализе новых логических схем, которые вводятся для описания этих процессов, будут обнаружены какие-то новые элементы, не используемые в формальном аппарате современной математики.

ОГЛАВЛЕНИЕ

УВАЖАЕМЫЙ ЧИТАТЕЛЬ!

Ваши замечания о содержании книги, ее оформлении, качестве перевода и другие просим присылать по адресу: 129820, Москва И-110, ГСП, 1-й Рижский пер., д. 2, издательство «Мир».

М. КАЦ, С. УЛАМ

Математика и логика
Ретроспектива и перспективы

Редактор *Г. М. Ильичева*
Художник *А. В. Шипов*
Художественный редактор *В. И. Шаповалов*
Технический редактор *З. И. Резник*
Корректор *Л. Д. Панова*

Сдано в набор 17/V 1971 г.
Подписано к печати 27/IX 1971 г.
Бумага № 3 $84 \times 108^1/_{32}$=4 бум. л. 13,44 усл. печ. л.
Уч.-изд. л. 11,55. Изд. № 1/6314
Цена 80 к. Зак. 1111

ИЗДАТЕЛЬСТВО «МИР»
Москва, 1-й Рижский пер., 2

Ордена Трудового Красного Знамени
Ленинградская типография № 2
имени Евгении Соколовой Главполиграфпрома
Комитета по печати при Совете Министров СССР
Измайловский проспект, 29