Математическое просвещение

У. У. СОЙЕР

ПРЕЛЮДИЯ К МАТЕМАТИКЕ

Перевод с английского М. Л. Смолянского и С. Л. Романовой

РАССКАЗ
о некоторых любопытных и удивительных областях математики
С ПРЕДВАРИТЕЛЬНЫМ АНАЛИЗОМ
математического склада
УМА
и целей математики

ИЗДАТЕЛЬСТВО «ПРОСВЕЩЕНИЕ»
МОСКВА 1972

Глава первая

О КРАСОТЕ И СИЛЕ

Нельзя быть настоящим математиком, не будучи немного поэтом.

К. Вейерштрасс

Мудрость рождается из пристального наблюдения за тем, как растут люди.

Конфуций

Эта книга о том, как растить математиков. Может быть, у вас и нет такого намерения. Но все же, я надеюсь, вы найдете в этой книге что-нибудь интересное для себя.

Сам, я, например, не собираюсь выращивать растения. Я никогда не занимаюсь садоводством, если могу обойтись без этого. Но я очень люблю смотреть на сады, выращенные другими. Мне было бы еще интереснее встретить человека, который объяснил бы (а очень мало садовников умеют это делать), как развивается растение; каким образом семечко в земле знает, куда направить стебель, а куда корни; как цветку удается поворачиваться всегда к солнцу; какие химические элементы растение получает из почвы и как ему удается превратить их в свою собственную, живую ткань. Интерес к этим вопросам совершенно не зависит от того, собираетесь ли вы действительно заниматься садоводством.

Я стараюсь писать эту книгу не с точки зрения садовода-практика, а для людей, которые хотят понять, что такое развитие. Я пишу не для учителей математики (хотя и учителя могли бы практически применить изложенные идеи), а для тех, кто хочет разобраться в характере мышления математика.

Иногда трудно передать то, что действительно достойно сообщения. Допустим, вы провели несколько лет в определенном месте; эти годы имеют для вас особое значение. Это могло быть в раннем детстве или в школьные годы, или же в период взрослой жизни, когда новые впечатления, приятные или неприятные, сделали вашу жизнь необычно интересной. Если вы вновь попадаете в это место, оно кажется вам каким-то особым. Ваши спутники, попавшие туда впервые, видят просто приятную деревеньку или обыкновенную городскую улицу. Они не видят того основного, что заставляет вас стремиться вновь посетить это место. Для того чтобы они поняли ваше стремление, вам нужно быть немного поэтом. Вы должны уметь описать это место так, чтобы передать ваши чувства.

Такое описание вполне возможно. Вообще, мы переоцениваем различия между людьми. Я уверен, что если бы кому-нибудь удалось на день стать кем-то другим, изменения были бы гораздо меньшими, чем можно ожидать. Чувства были бы теми же самыми, но обращенными на другие предметы!

Вообще говоря, в процессе обучения передаются скорее сведения о предметах, чем живой ход мысли. Допустим, например, что ко мне приходит человек с каким-либо вопросом: это может быть неясная задача из школьной арифметики или серьезная научная проблема. Допустим, мне удается решить эту задачу; тогда мне очень легко объяснить ее решение. Представим себе, что я так и сделал, т. е. что я объяснил этому человеку, как поступать в данном конкретном случае. Но если он натолкнется на задачу другого типа, он опять обратится ко мне, так как я сообщил ему только решение данной задачи, но не обучил его самостоятельному мышлению. Я бы почувствовал настоящее удовлетворение лишь в том случае, если бы смог передать моему ученику не просто знания, а гибкость ума, которая дала бы ему возможность в дальнейшем самостоятельно решать задачи.

Само собой разумеется, здесь существуют определенные пределы, о которых не следует забывать. Ум — это один из факторов, важных для решения проблем, а он часто бывает врожденным. Но, кроме ума, имеется ряд факторов, которые зависят главным образом от образования и воспитания, а именно, чувство страха или уверенности, привычка полагаться на себя, инициатива и настойчивость. Я не думаю, чтобы мы сильно отличались от наших пещерных предков по врожденным качествам ума. Все историческое развитие до настоящего времени, все государственные различия между отдельными странами существенно связаны с изменениями в системе образования [1].

Книжки, подобные этой, сами по себе свидетельствуют о серьезных переменах. Ни в прошлом веке, ни тем более раньше не было такого громадного количества читающих и думающих людей. А сейчас они образуют своего рода незримое братство, интернациональный университет.

В настоящее время, когда наши познания природы вещей столь велики, а наше понимание самих себя столь мало, становится чрезвычайно важной правильная оценка огромных, неиспользованных возможностей образования. Промышленная революция предполагает и требует революции в психике (которая очень мало изменилась с тех времен, когда люди отказывались верить, что паровозы могут двигаться).

Настоящая книга написана в стране, где происходили и происходят огромные перемены в системе образования. В 1948 году был основан университет Золотого Берега [2]. Для поступления в универ-

[1] Клакхон (Kluckhohn) в своей книге «Mirror for Man» приводит интересный пример американского мальчика, воспитанного в Китае. Он по внешнему виду остался американцем, а во всех других отношениях стал похож на китайца.

[2] Золотой Берег — прежнее название республики Гана.— *Прим. ред.*

ситет отбирались лишь талантливые, любознательные молодые люди. Исходя из возможностей их врожденного ума, они все через десять — двадцать лет должны были стать научными работниками в области математики или профессорами университета. Но в это время в стране не существовало математических традиций. Их следовало создать.

Но для этого нам надо было как следует разобраться в существе встающих перед нами задач. Надо было изучить жизнь будущего математика в одном из старых математических центров; рассмотреть все те факторы, которые помогают ему развиваться: атмосферу школы и колледжа, бесчисленные идеи, намеки и указания, которые он получает от преподавателей и из книг. Из всего этого нам следовало составить для себя ясное представление о том, что мы пытаемся сделать, какими качествами должен обладать математик и как их необходимо развивать.

В первых пяти главах я как раз пытаюсь определить, какими характерными качествами должен обладать математик и как он развивается. Эти главы содержат также ряд математических этюдов, предназначенных для иллюстрации того, что интересует математика. В последующих главах излагаются различные разделы математики, отобранные мною по признаку их необычности, новизны и богатству применений. Все это вполне элементарные теории; чтобы читатель мог разобраться в них, ему достаточно самых смутных воспоминаний о школьной математике.

К расчетам я прибегал редко и нигде не использовал их в качестве составной части доказательства. Этот факт интересен: он показывает, что целый ряд разделов современной математики (с 1800 г.) не является развитием старых работ, а идет в совершенно новом направлении.

Вообще, вы не найдете здесь длинных математических выкладок. Почти все математические открытия имеют в основе очень простую идею. Учебники часто скрывают этот факт. Они обычно содержат громоздкие выводы и этим создают впечатление, что математики —

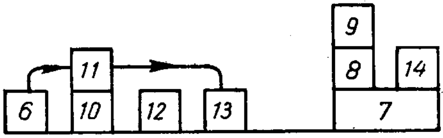


Рис. 1. Схема построения книги.

Главы 1—5, использующие математику только для иллюстрации основных качеств, которыми должен обладать математик, на схеме не указаны. Расположение на схеме одной главы под другой показывает, что в верхней главе используются идеи, изложенные в нижней; так, чтобы понять главу 11, необходимо прочесть главу 10.

Стрелка от одной главы к другой указывает на существующую между ними связь. Например, один из разделов главы 13 («Конечные геометрии») не может быть понят без ссылок на материал главы 11. Все же остальные разделы главы 13 можно читать так, как если бы это была первая глава книги. При чтении книги вы убедитесь, что одноэтажные дома встречаются в ней гораздо чаще, чем небоскребы.

это люди, которые всю жизнь просиживают за письменными столами и переводят тонны бумаги. Это чепуха. Многие математики очень успешно работают в ванной, в кровати, ожидая поезда или катаясь на велосипеде (предпочтительно при слабом уличном движении). Математические вычисления производятся до или после открытия. Само открытие возникает из основных идей. Именно эти основные идеи я и намереваюсь изложить. Конечно, нельзя обойтись и без некоторых деталей, если мы хотим, чтобы это была книга по математике, а не просто сентиментальная рапсодия в честь математики.

Но прежде чем вы расстанетесь со своими деньгами, покупая эту книгу, я хочу честно предупредить вас, что в некоторых местах вам все же встретятся довольно длинные алгебраические выражения. Один из вопросов, которые я как раз хочу обсудить, это — как нужно рассматривать алгебраическое выражение. Чем длиннее выражение, тем важнее знать, как его рассматривать. Например, в главе 9, посвященной определителям, вы увидите довольно громоздкие группы символов. Они приведены для того, чтобы мы могли воскликнуть: «Ну и путаница! Как же в этом хаосе можно найти ту простую идею, которую мы собираемся рассмотреть. Где та стройность формы и внутренняя закономерность, без которых немыслима математика?»

Следует помнить, что в каком-то смысле высшая математика проще элементарной. Исследовать, например, лесную чащу пешком очень трудно, с самолета это делается проще.

Подчеркиваю еще раз, что эта книга не ставит себе целью дать обзор развития математики с 1800 года, а также не является кратким сообщением о том, чем занимаются математики в настоящее время.

В книге просто рассматриваются отдельные вопросы из области математики, не более. Сомневаюсь, чтобы можно было сделать больше в книге подобного объема.

СФЕРА МАТЕМАТИКИ

Немногие представляют себе, как огромна сфера действия современной математики. Вероятно, было бы легче овладеть всеми существующими языками, чем всеми математическими знаниями, известными в настоящее время. Мне кажется, что все языки можно было бы выучить за одну человеческую жизнь; а всю математику, конечно, нет. К тому же объем математических знаний не остается неизменным. Ежегодно публикуются все новые открытия. Например, в 1951 году для реферативного изложения всех математических статей, вышедших за год, потребовалось 900 печатных страниц крупного формата. Только за январь упомянуто 451 название, причем реферировались статьи и книги, рассматривающие новые проблемы; лишь в немногих из них упоминались известные факты.

Человеку, желающему быть в курсе всего нового в математике, пришлось бы прочитывать ежедневно около 15 статей, весьма больших по объему и содержащих сложные математические выкладки. Трудно даже мечтать о выполнении подобной задачи.

Открытия, которые делают математики, столь разнообразны по своему характеру, что однажды кто-то, видимо, в отчаянии предложил определить математику как «все, чем занимаются математики». Казалось, что только такое широкое определение может охватить все, что относится к математике. Математики решают проблемы, которые в прошлом не считались математическими, и трудно предсказать, чем они еще будут заниматься в будущем.

Точнее было бы определение: «Математика — это классификация всех возможных задач и методов их решения». Это определение, пожалуй, тоже расплывчато, так как оно охватывало бы даже такие рубрики, как газетные объявления «Обращайтесь со всеми вашими сердечными заботами к тете Минни», что мы никак не имеем в виду.

Для целей нашей книги достаточно было бы определение: «Математика — это классификация и изучение всех возможных закономерностей». Слово «закономерность» здесь используется в таком смысле, с которым многие могут не согласиться; а именно, в самом широком смысле, как название *любого рода закономерностей, которые могут быть познаны умом*. Жизнь, особенно интеллектуальная жизнь, возможна лишь потому, что в мире существуют определенные закономерности [1]. Птица, например, различает черные и желтые полоски осы; человек знает, что растение развивается из посаженного в землю семени. В обоих случаях существует осознанная закономерность.

Закономерность — это наиболее стабильная характеристика постоянно меняющегося мира. Сегодняшний день не может быть похож на вчерашний. Нельзя увидеть дважды одно и то же лицо под одним и тем же углом зрения. Узнавание возможно не потому, что опыт точно повторяется, а потому, что в огромном разнообразии жизненных явлений можно распознать определенные закономерности. Такую постоянную закономерность мы имеем в виду, когда говорим «мой велосипед» или «река Темза», хотя велосипед довольно быстро изнашивается, а воды реки непрерывно переливаются в море.

Любая математическая теория должна непременно сочетать в себе мощь метода, обусловливающую возможность применений к естественным наукам, и красоту, стройность, столь привлекательную для ума. Нам кажется, что наше определение математики удовлетворяет обоим этим требованиям. Вся наука зиждется на закономерностях, имеющихся в природе; для практики важна классификация этих закономерностей. С другой стороны, ум должен получать

[1] Ср. «Наука и метод» Пуанкаре.

удовольствие при изучении и познании закономерностей, поскольку необходимость и желание всегда связаны в природе. Если следование законам природы характерно как для человека, так и для животных, то естественно ожидать, что подчинение закономерностям доставляет удовольствие, такое же, как удовлетворение любой другой человеческой потребности.

Интересно заметить, что «чистые» математики, движимые только чувством стройности и математической формы, часто приходили к выводам, которые в дальнейшем оказывались чрезвычайно важными для науки. Греки изучали свойства эллипса более чем за тысячу лет до того, как Кеплер использовал их идею для определения траекторий планет. Математический аппарат теории относительности был создан за тридцать — пятьдесят лет до того, как Эйнштейн нашел для него применение в физике. Подобных примеров можно было бы привести много. С другой стороны, много стройных теорий и проблем, которые любой «чистый» математик причислит к математике, возникли в связи с физикой.

КАКИЕ ЗАКОНОМЕРНОСТИ ЛЮБИТ ПРИРОДА?

Еще один интересный факт. Мы иногда встречаем одну и ту же закономерность в разных областях природы, как если бы число закономерностей было крайне лимитировано. Закономерность, которую математики обозначают Δv, встречается более чем в десятке различных областей науки. Она возникает в связи с такими явлениями, как тяготение, свет, звук, теплота, магнетизм, электростатика, электрический ток, электромагнитные излучения, морские волны, полет самолета, колебания упругих тел и строение атома, не говоря уже об одной чисто математической теории первостепенной важности — теории функций комплексного переменного.

Практики часто допускают ошибку, рассматривая все эти проявления Δv по отдельности. Это приводит к излишней трате сил. Мы здесь имеем дело не с двенадцатью отдельными теориями, а с одной теорией, имеющей 12 применений. Здесь везде проявляется одна и та же закономерность. Физически ее применения различны, математически — одинаковы.

Идея наличия одной и той же закономерности в различных условиях очевидна. Остается только придумать для нее соответствующее название; и вот мы имеем один из самых общих терминов современной математики — *изоморфный* (ιςος — подобный, μορφη — форма), т. е. «имеющий одну и ту же форму». Ничто не доставляет математику большего наслаждения, чем открытие, что две вещи, которые он ранее считал совершенно различными, оказываются математически идентичными, изоморфными. «Математика,— говорит Пуанкаре,— это искусство называть разные вещи одним и тем же именем».

Возникает вопрос: «Чем вызвано то, что эта закономерность встречается так часто?» Здесь мы уже лавируем на грани математического мистицизма. Окончательного ответа на этот вопрос быть не может. Предположим, мы доказали, что эта закономерность имеет ряд свойств, делающих ее особенно подходящей; но тогда бы неминуемо возник следующий вопрос: «Почему же природа предпочитает именно эти свойства?» — и так без конца. Тем не менее на вопрос, почему именно закономерность Δv встречается так часто, можно частично ответить [1].

Невозможность дать окончательный ответ на вопрос «Почему Мир устроен так, а не иначе?» вовсе не означает, что такая постановка вопроса абсолютно бесполезна. Нам, может быть, удастся показать в будущем, что все научные законы, открытые до сих пор, имеют ряд общих свойств. Математик, изучающий закономерности этих общих свойств, имеет все основания надеяться, что его работа окажется полезной для будущих поколений; хотя, конечно, абсолютной уверенности в этом быть не может — ни в чем нельзя быть абсолютно уверенным. Кроме того, он может надеяться удовлетворить таким образом свою собственную потребность в глубоком проникновении в законы Вселенной.

МАТЕМАТИК КАК КОНСУЛЬТАНТ

Практики, как правило, не имеют представления о математике как о способе классификации всех проблем. Обычно они стремятся изучать только те разделы математики, которые уже оказались полезными для их специальности. Поэтому они совершенно беспомощны перед новыми задачами. Вот тогда-то они и обращаются за помощью к математикам. (Это разделение труда между инженерами и математиками, вероятно, оправдано; жизнь слишком коротка для того, чтобы одновременно изучать и абстрактную теорию и инженерное дело.) Встреча математика и инженера обычно очень забавна. Инженер, ежедневно имея дело с машинами, настолько привыкает к ним, что не может понять чувств человека, видящего машину впервые. Он забрасывает своего консультанта-математика огромным количеством подробностей, которые для того ровным счетом ничего не значат. Через некоторое время инженер приходит к выводу, что математик — абсолютный невежда и что ему нужно объяснять простейшие вещи, как ребенку или Сократу. Но как только математик поймет, что делает машина или что от нее требуется, он переводит

[1] Объяснение может заключаться в том, что в вакууме все точки пространства равноценны, а также нет преимущественных направлений. Поэтому нельзя ожидать, чтобы законы, справедливые в вакууме, выделяли какую-нибудь отдельную точку или направление. Это значительно ограничивает выбор возможных законов. Закономерность $\Delta v=0$ выражает символически тот факт, что величина v в любой точке равна среднему значению v на поверхности шара с центром в этой точке. Для этого закона все точки и все направления одинаковы. Это простейший закон такого рода.

задачу на язык математических терминов. После этого он может заявить инженеру одно из трех:

1) что задача известна и уже решена;

2) что это новая задача, которую он может попытаться решить;

3) что это одна из тех задач, которую математики безуспешно пытались решить, и что еще могут пройти века, прежде чем будет сделан хотя бы шаг к ее решению, и что поэтому инженеру придется решать ее эмпирически.

К сожалению, третий случай встречается удручающе часто. Но первый и второй случаи также довольно часты, и вот тогда-то математик, благодаря его знанию закономерностей, может принести пользу в тех областях, о которых он в некотором смысле ничего не знает.

Математик, желающий консультировать инженеров, должен поэтому изучать не только те задачи, которые уже встречались; он также должен быть готов анализировать задачи, которые только еще могут возникнуть. Можно считать, что практические задачи образуют в известном смысле определенный тип задач. Например, очень часто практическая задача принимает форму дифференциального уравнения. Некоторые дифференциальные уравнения мы умеем решать, а другие нет. Поэтому математик стремится расширить свой арсенал, изучая те дифференциальные уравнения, которые пока еще не решены; это неизбежно поставит перед ним ряд фундаментальных вопросов, например: «Какая разница между уже решенными и нерешенными дифференциальными уравнениями? В чем состоит трудность их решения?»

Но не всегда практическую задачу удается отнести к какому-либо математическому типу. Иногда возникают задачи, совершенно не похожие на уже встречавшиеся. Ключ к их решению может быть найден совершенно случайно. Они даже могут напомнить какую-нибудь задачку, решенную на досуге. В подобном случае такая задача может лечь в основу новой крупной теории. Впрочем, эта доктрина, как и все доктрины, может привести к заблуждениям. Человек может потратить всю свою жизнь на решение пустяковых задач, теша себя надеждой, что они, возможно, станут началом новых областей математики. И так, конечно, может случиться; все зависит от умения определить, что может оказаться важным, и нет правила, которое позволило бы судить о правильности выбора. Любой математик согласится, что существуют теории, которые до сих пор не нашли практического применения, но чувствуется, что они являются очень важной частью математики. Работа над ними свидетельствует не об отходе от главного направления, а как раз об обратном — о борьбе за развитие математики. Когда-нибудь эти теории, «подобно эллипсу», найдут своего Кеплера и, «подобно тензорному анализу», своего Эйнштейна. Но, во всяком случае, ими подготовлен мощный аппарат для решения определенного класса задач, если возникнет такая необходимость.

МАТЕМАТИК КАК ХУДОЖНИК

Я пишу и представляю себе, как «чистый математик» читает эту книгу со все возрастающим недовольством. Даже если предположить, что у него хватило терпения дочитать до сих пор, он, наверное, говорит: «Вы рассматриваете математику как нечто приносящее практическую пользу. Но математика не прикладная наука, она важна сама по себе. Важна не полезность математики, а ее стройность. Прикладная математика — это самая скучная часть математики. Посмотрите-ка на людей, работающих над теорией чисел, не имеющей никакого практического применения,— вы предложили бы им заниматься бухгалтерией?»

Этой точке зрения, которой придерживаются многие крупные математики, можно противопоставить утилитарный, бюрократический взгляд на математику. Согласно этой немного пуританской математики должны стыдиться своей приверженности к красоте и стройности, и им следует заниматься лишь решением неотложных практических задач, получаемых от практиков.

Обе точки зрения несовершенны. Ясно, что следование любой из них до логического конца явилось бы роковым и для математики, и для технического прогресса.

Давайте рассмотрим вначале точку зрения «математика ради математики» и проанализируем как частный пример ее применение к положению в стране Золотой Берег. Золотой Берег — это страна, где живут умные, жизнерадостные люди. Не надо думать, что это несчастная страна, но все же это страна, которая нуждается во многом. Во многих местах плохое водоснабжение, отсутствует канализация, распространены заразные болезни, не хватает продуктов питания, некоторые дети хронически недоедают. Так как же в этих условиях мы можем оправдать расходы математического факультета нового университета? Ссылками на стройность математики? В таких условиях защищать расходы на факультет математики только на основании прелести математики было бы верхом бессердечия. Математика имеет культурную ценность, но ее ценность состоит вовсе не в том, чтобы любоваться новыми закономерностями, с полным безразличием ко всему окружающему. Совершенно очевидно, что практическая сила математики как прикладной науки для техники, биологии и медицины имеет первостепенное значение в любой развивающейся стране. Красота без силы — бессмысленна.

Следует также сказать несколько слов в защиту математика-художника. Полезность без красоты бессильна. Работа, которой занимаются только ради ее результата, без какого бы то ни было удовольствия от самого процесса работы, может оказаться выполненной плохо. Инженер иногда начинает изучать математику потому, что она полезна для него как для специалиста, но если он больше ничем не интересуется, если он не увлечется самим предметом, математика мало ему поможет.

Утилитарный подход к математике оправдан лишь в одном отношении; он признает тот факт, что математик — человек, что он зависит от усилий других людей: ему нужна пища, одежда, кров и очаг — и за все это он должен как-то отплатить обществу. Такую постановку вопроса легче всего принимают нематематики: администраторы и налогоплательщики. Вообще, соображения полезности приводят к заключению, что математики нужны; но они не показывают путей, как их создавать. Было бы весьма желательным, чтобы Сахара покрылась дубовыми рощами, в тени которых отдыхали бы путешественники, но одного желания мало, чтобы там выросли деревья. Математикам, как и деревьям, нужны условия, пригодные для роста.

Можно, конечно, возразить, что люди не деревья, что если человек понимает, что нужно что-то сделать, он в состоянии приняться за это. Эта мысль справедлива лишь в определенных пределах. Социальные условия в какой-нибудь стране могут быть благоприятны для развития математики; если стране срочно нужны математики и все об этом знают, тогда математика может там процветать. Но это все же не отвечает на вопрос, как начинается это процветание. Внешних побудительных факторов, будь то хороших или плохих, еще недостаточно. Такие чувства, как жажда денег, стремление к славе или любовь к человечеству, не могут сделать из человека великого композитора. Многим молодым людям хотелось бы сидеть за пианино и импровизировать перед восторженной толпой. Но многие ли способны на это? — Желать — еще не значит уметь. Чтобы сочинять музыку, нужно любить музыку, а не успех (или, во всяком случае, музыку не меньше, чем успех). А чтобы стать математиком, нужно увлекаться прелестью закономерностей и логической стройностью законов. Это не значит, конечно, что такое увлечение должно быть единственной страстью, у вас могут быть и другие цели, вы можете заниматься другими делами, но если вы не попадете под очарование математики, вы в математике ничего не совершите.

В этом смысле точка зрения математика-художника ближе к жизни, чем отношение математика-бюрократа. Первый, по крайней мере, понимает, как люди становятся математиками. Как чистый художник, так и чистый бюрократ ошибаются или, во всяком случае, не совсем правы. Если бы обучение математике должно было основываться лишь на одной из этих теорий, то можно сказать с уверенностью, что теория художника принесла бы гораздо меньше вреда. Художник может быть анархистом, представителем богемы или даже бродягой, но он, по крайней мере, живой человек, а без жизни нет роста. И если такой человек научит детей любить свой предмет, то всегда остается надежда, что его ученики в зрелом возрасте обратят свои способности на полезные дела. Но если вы отдадите их в руки человека с крайне утилитарными взглядами, то они, наверное, не принесут никакой пользы обществу, так как растеряют все свои способности.

Глава вторая

КАКИМИ КАЧЕСТВАМИ ДОЛЖЕН ОБЛАДАТЬ МАТЕМАТИК?

Не представляю себе, как можно довольствоваться знаниями, полученными из вторых рук; хотя чужое знание может нас кое-чему научить, мудр бываешь лишь своей собственной мудростью.

Монтень, О педантизме

Для всех математиков, характерна *дерзость ума*. Математик не любит, когда ему о чем-нибудь рассказывают, он сам хочет дойти до всего. Конечно, зрелый математик, узнав о каком-нибудь великом открытии, поинтересуется, в чем оно состоит, и не станет терять время на то, чтобы открывать уже открытое. Но я имею в виду юных математиков, у которых дерзость ума проявляется особенно сильно. Если вы, например, преподаете геометрию девяти-десятилетним ребятам и рассказываете им, что никто еще не смог разделить угол на три равные части при помощи линейки и циркуля, вы непременно увидите, что один-два мальчика останутся после уроков и будут пытаться найти решение. То обстоятельство, что в течение двух тысяч лет никто не решил эту задачу, не помешает им надеяться, что они смогут это сделать в течение часового перерыва на обед. Это, конечно, не очень скромно, но и не свидетельствует об их самонадеянности. Они просто готовы принять любой вызов. А ведь в действительности уже доказано, что невозможно разделить угол на три равные части при помощи линейки и циркуля. Их попытка найти решение — того же рода, что попытка представить $\sqrt{2}$ в виде рациональной дроби p/q.

Хороший ученик всегда старается забежать вперед. Если вы ему объясните, как решать квадратное уравнение дополнением до полного квадрата, он непременно захочет узнать, можно ли решить кубическое уравнение дополнением до полного куба. Остальные ученики класса не задают подобных вопросов. С них хватит и квадратных уравнений, они не ищут дополнительных трудностей.

Вот это *желание исследовать* является второй отличительной чертой математика. Это одна из сил, содействующих росту математика. Математик получает удовольствие от знаний, которыми уже овладел, и всегда стремится к новым знаниям.

Эту мысль можно пояснить на примере дробных показателей степени из школьного курса алгебры. Легко представить себе человека, который, поверхностно познакомившись с дробными и отрицательными показателями степени, начнет недоумевать, зачем все это нужно. Ведь приходится преодолевать столько логических трудностей! Мне представляется, что тот, кто открыл дробные показатели степени, сначала работал над целыми показателями и получил такое большое удовлетворение от этой работы, что ему захотелось развить этот раздел, и он готов был взять на себя логический риск. Ведь на первых порах новое открытие почти всегда является вопросом веры и лишь позднее, когда становится ясным, что это действительно открытие, приходится находить логическое оправдание, которое удовлетворит самых придирчивых критиков.

Я уже говорил об *интересе к закономерностям* — третьем необходимом качестве математика. Уже в самом начале арифметики встречаются закономерности. Например, из четырех одинаковых камней можно сложить квадрат, а из пяти — нельзя. Математические, как и музыкальные, способности проявляются очень рано, с четырех лет, а иногда и раньше. Один малыш однажды сказал мне: «Мне нравится слово September [сентябрь], ведь получается sEptEmbEr». Сам я никогда не замечал закономерного чередования гласных и согласных в этом слове. Оно действительно совершенно симметрично. Такому ребенку, конечно, понравится арифметика.

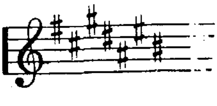
В таблице умножения имеется один элементарный пример закономерности. Обычно дети любят умножать на 2 и на 5, потому что последние цифры ответа легко запомнить: при умножении на 2 всегда получаются четные цифры, а при умножении на 5 еще проще: всегда 0 или 5. Но даже в умножении на 7 есть свои закономерности. Если мы посмотрим на последние цифры произведений 7, 14, 21, 28, 35, 42, 49, 56, 63, 70, т. е. на 7, 4, 1, 8, 5, 2, 9, 6, 3, 0, то увидим, что разность между последующей и предыдущей цифрами составляет: -3, $+7$; -3, -3, $+7$; -3, -3, -3. В этом ряду чувствуется совершенно определенный ритм [1]. Если прочесть последние цифры ответов при умножении на 7 в обратном порядке, то мы получаем последние цифры от умножения на 3.

Даже в начальной школе можно развить навык наблюдения за математическими закономерностями. Бóльшая часть ранних работ Гаусса явилась следствием его привычки делать вычисления и анализировать полученные результаты. Эрмит, один из крупнейших французских математиков, также подчеркивал, как часто наблюдательность ведет к математическим открытиям. Правда, одной наблюдательности мало, чтобы стать великим математиком.

Так же как умение замечать арифметические закономерности помогает школьнику решать арифметические примеры, умение наблюдать закономерности в алгебре помогает избегать ошибок или обнаруживать описки.

---

[1] Подобная закономерность встречается в записи диезов и бемолей при ключе.



Запись семи диезов представляет собой последовательность шагов: 3 вниз, 4 вверх, 3 вниз, 4 вверх, 3 вниз.

Например, условием равенства корней квадратного уравнения вида $ax^2+2bx+c=0$ является

$$b^2-ac=0.$$

(Читателю, возможно, более знакомо условие $b^2=4ac$ для уравнения $ax^2+bx+c=0$.) Подобное же условие можно найти для равенства корней кубического уравнения. Кубическое уравнение имеет три корня; спрашивается, в каком случае два из них равны.

Для уравнения $ax^3+3bx^2+3cx+d=0$ условием равенства двух корней будет $(bc-ad)^2-4(ac-b^2)(bd-c^2)=0$. Это условие можно также записать в виде

$$a^2d^2-6abcd+4b^3d+4ac^3-3b^2c^2=0.$$

Если вы проанализируете левые части указанных условий, вы обязательно заметите следующие закономерности:

(1) Общее количество множителей в каждом члене выражения одинаково. Например, в выражении b^2-ac каждый член состоит из двух букв, умноженных друг на друга, т. е. степень каждого члена равна двум. В более длинном условии для равенства двух корней кубического уравнения каждый член содержит четыре буквы, умноженные друг на друга; например, b^3d есть сокращенное $bbbd$.

(2) В этих выражениях имеется также равновесие другого рода, хотя эта вторая закономерность не сразу заметна: равновесие между буквами, которые стоят раньше или позже в алфавите. Например, в условии для кубического уравнения имеются члены $abcd$ и $a^2d^2=aadd$. Сравним их вторые буквы. В члене $aadd$ вторая буква есть a, в члене $abcd$ вторая буква есть b. Буква a стоит в алфавите раньше , чем b. Но справедливость торжествует: если мы посмотрим на третьи буквы этих выражений, то видим, что $aadd$ содержит d, в то время как $abcd$ содержит c; буква d стоит в алфавите позже c. Равновесие совершенно точное: a стоит перед b, а d после c. То же самое равновесие сохраняется во всех членах. Это можно проверить следующим образом. Оценим каждую из букв a, b, c, d определенным количеством очков: $a=0$, $b=1$, $c=2$, $d=3$. Тогда сумма очков в каждом члене будет равна 6; например, для $abcd$ имеем $0+1+2+3=6$, для ac^3 $0+2+2+2=6$.

Число очков, сопоставленное данной букве, называют ее весом; таким образом, вес каждого члена равен 6. В выражении b^2-ac вес каждого члена равен 2.

(3) Сумма численных коэффициентов в каждом выражении равна нулю. Действительно, в b^2-ac коэффициентами являются числа $+1$ и -1, сумма которых равна нулю, а в выражении

$$a^2d^2-6abcd+4b^3d+4ac^3-3b^2c^2$$

—числа 1, —6, 4, 4, —3, которые в сумме также дают 0. Другими словами, если в приведенных выражениях положить $a=b=c=d=1$, то эти выражения обратятся тождественно в нуль.

Эти три свойства дают возможность проверить нашу внимательность и аккуратность. Свойство (1) позволяет проверить, не ошиблись ли мы при написании показателей степеней букв, поскольку из-за такого рода ошибки количество букв в различных членах может оказаться неодинаковым (если, конечно, мы не допустили две взаимно компенсирующиеся ошибки). Свойство (2) дает возможность избежать ошибки при переписывании буквенных символов. Если, например, на каком-то этапе работы мы написали d вместо a, это изменит вес данного члена. Свойство (3) спасает нас от ошибок при сложении членов. Положим, мы пропустили один член при сложении. Тогда, придав всем буквам значение 1, мы не получим в сумме 0 и тем самым сразу обратим внимание на ошибку.

Такой тип контроля не дает еще полного представления о математике. Посмотреть на ответ и проверить, получена ли ожидаемая симметрия или равновесие,— это одна из вещей, которые математик делает почти бессознательно. Просто удивительно, как редко детей учат анализировать именно так. То и дело на экзаменах ученики сдают работы, в которых ответы имеют «какую-то странную форму»,— как любил выражаться отец Браун у Честертона [1]. Применения этого типа контроля гораздо шире, чем специальный вид проверки, о котором мы говорили.

ПРИМЕР ГЕОМЕТРИЧЕСКОЙ ЗАКОНОМЕРНОСТИ

В школьном курсе геометрии встречается уравнение

$$x^2 = \frac{(ac+bd)(ad+bc)}{ab+cd},$$

где x, a, b, c и d обозначают длины отрезков на рис. 2.

Этот результат легко доказать, выразив $\cos Q$ и $\cos S$ через стороны треугольников PQR и PSR. Сумма косинусов углов Q и S должна быть равна 0, поскольку $Q+S=180°$ (вписанный четырехугольник), и поэтому $\cos S = -\cos Q$. Полученное уравнение затем разрешается относительно x^2.

В этом доказательстве нет ничего интересного, но примечательна закономерность результата.

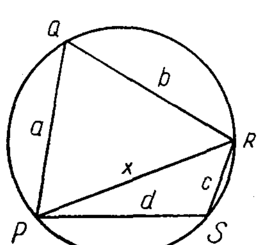


Рис. 2.

Четыре предмета можно разбить попарно тремя различными способами. Так, если четверо играют в бридж, то A и B могут играть против C и D, или A и C против B и D или же A и D против B и C. Никак иначе сгруппировать нельзя.

[1] Герой рассказов английского писателя Г. К. Честертона священник Браун раскрывал преступления, подмечая самые незначительные отклонения от нормы.— *Прим. ред.*

Алгебраические выражения $ab+cd$, $ac+bd$, $ad+bc$ также составлены на основе группировки символов a, b, c, d в пары и последующего сложения этих пар. Таким путем можно составить три, и только три, выражения (считая, конечно, что $cd+ab$ — это то же самое, что $ab+cd$).

В приведенной выше формуле встречаются все три выражения: два — в числителе, третье — в знаменателе.

Я не собираюсь здесь вдаваться в вопрос, почему это получается именно так. Мне хочется только подчеркнуть, что формула благодаря такой закономерности легко запоминается.

СМЫСЛ И ОБОБЩЕНИЕ

Иногда, рассматривая какое-нибудь произведение искусства, мы восхищаемся его красотой и чувствуем его значительность, но не можем сказать, в чем они состоят. Лучше и не пытаться это делать. Один поэт протестовал против варварского обычая заставлять детей пересказывать стихи своими словами. Он говорил, что единственный способ объяснить содержание стихотворения — это написать лучшее стихотворение. А от детей этого, конечно, требовать нельзя.

Но в математике, как правило, бывает иначе. Если нам встречается закономерность, мы обязательно спрашиваем: «Почему она встречается? Что она означает?» И мы обычно можем найти ответы на эти вопросы. Итак, всякий раз, когда появляется закономерность, математик чувствует, что он обязан узнать, почему она появляется.

Мы можем объяснить, например, почему условие равенства корней должно обладать указанными выше тремя свойствами. Я лишь намечу общую идею, опустив алгебраические подробности (которые в действительности совсем просты).

Рассмотрим кубическое уравнение, которое я для краткости напишу в виде $f(x)=0$. Первый график, показанный на рис. 3*а*, представляет функцию $y=f(x)$ в случае, когда корни уравнения — действительные и различные. Если эту кривую сместить вверх так, чтобы точки B и C совпали, мы получим вторую кривую, изображенную на рис. 3*б*, которая соответствует кубическому уравнению с двумя корнями. Между прочим, мы можем легко проверить равенство корней при помощи вычисления. В точке E величины y и $\frac{dy}{dx}$ равны 0. Известно, что равные корни получаются в том случае, когда уравнение $\frac{dy}{dx}=f'(x)=0$ имеет общий корень с исходным уравнением $f(x)=0$.

Теперь представьте себе, что вторая кривая (рис. 3*б*) нарисована на куске резины. Предположим, что резину растягивают вертикально. Уравнение кривой приняло бы тогда вид $y=kf(x)$. Очевидно, что если исходная кривая касалась оси x (как это должно быть

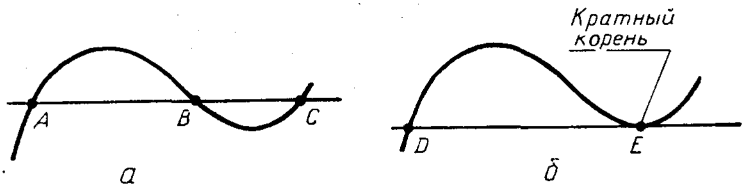


Рис. 3.

при равных корнях), то новая кривая, полученная при вертикальном растяжении, также бы ее касалась. Иначе говоря, если уравнение $f(x)=0$ имеет равные корни, то это же справедливо и для уравнения $kf(x)=0$. Рассматривая вопрос о влиянии множителя k на константы a, b, c и d, т. е. на коэффициенты многочлена $f(x)$, нетрудно видеть, что необходимым условием равенства корней является свойство (1) — все члены должны иметь одну и ту же степень.

Точно так же не повлияет на общий вид кривой растяжение по оси абсцисс. Она будет по-прежнему касаться оси x. Отсюда мы можем заключить, что если уравнение $f(x)=0$ имеет равные корни, то и $f(kx)=0$ тоже имеет равные корни. Это приводит нас к свойству (2) — все члены должны иметь одинаковый вес.

Свойство (3) оказывается самым простым. Если мы положим $a=b=c=d=1$, то квадратное уравнение примет вид $x^3+2x+1=0$, а кубическое $x^3+3x^2+3x+1=0$. Эти уравнения короче можно записать как $(x+1)^2=0$ и $(x+1)^3=0$ соответственно; отсюда очевидно, что они имеют повторяющийся корень —1. (Все три корня кубического уравнения равны —1; для наших целей было бы достаточно, если бы только два из них были равны —1.)

Таким образом, свойства (1) *и* (2) *справедливы для любого условия, которое сохраняется при вертикальном и горизонтальном растяжениях (т. е. при изменении масштабов по осям y и x соответственно*[1]*); свойство* (3) *справедливо для любого условия, которое выполняется для уравнения* $(x+1)^n=0$.

Эти выводы являются большим шагом вперед по сравнению с тем моментом, когда мы просто *подметили* свойства (1), (2) и (3). Мы их интерпретировали и знаем теперь, когда и почему они сохраняются.

Все это позволяет нам применить свойства (1), (2) и (3) не только для проверки условия равенства корней, но и для других целей. Например, если мы имеем кубическое уравнение с тремя корнями, мы можем исследовать, при каких условиях один корень окажется точно посредине между двумя другими. Для соответствующей кривой (рис. 3*а*) это означало бы, что точка B делит отрезок AC попо-

[1] Примером такого условия является рассмотренное выше условие равенства корней.— *Прим. ред.*

лам. Это свойство не нарушалось бы при изменении масштабов как по вертикали, так и по горизонтали. Оно справедливо и для уравнения $(x+1)^3=0$, потому что в этом случае точки A, B и C совпадают с $x=-1$ и, таким образом, B оказывается серединой отрезка AC. Следовательно, алгебраическое выражение этого условия должно обладать свойствами (1), (2) и (3); и действительно, оно имеет ожидаемый вид:

$$2b^3-3abc+a^2d=0.$$

Это пример *обобщения* — одного из самых важных факторов развития математики. Мы начали от наблюдения, касавшегося только условия равенства корней, а закончили принципом, который применим к гораздо более широкому классу условий. Ценность такого обобщения очевидна. Чем шире круг вопросов, к которым применим какой-нибудь общий принцип, тем чаще он нам поможет выпутаться из затруднений. Пуанкаре говорил: «Предположим, я занялся сложным вычислением и с большим трудом, наконец, получил результат; но все мои усилия окажутся напрасными, если они не помогут предвидеть результат в других аналогичных вычислениях, если они мне не дадут возможность проводить их с уверенностью, избегая тех ошибок и заблуждений, с которыми я должен был мириться в первый раз» [1].

ОБОБЩЕНИЕ И ПРОСТОТА

После обобщения результат становится более полезным. Вас, возможно, удивит, что обобщение почти всегда также упрощает результат. Более общий вывод легче воспринять, чем менее общий.

Иллюстрацией может послужить следующая тривиальная задача: «В одном стакане десять ложек воды, в другом — десять ложек вина. Ложку воды из первого стакана переливают во второй. Смесь энергично взбалтывается, затем ложку полученной смеси переливают в первый стакан. Спрашивается, чего будет больше: вина в первом стакане или воды во втором?».

Очевидно, для того чтобы ответить на этот вопрос, можно проделать следующее вычисление. После того как ложку воды перелили во второй стакан, там стало 10 ложек вина и 1 ложка воды, т. е. всего 11 ложек жидкости. Таким образом, одна ложка смеси содержит $\frac{10}{11}$ ложки вина и $\frac{1}{11}$ ложки воды. Переливаем одну ложку смеси в первый стакан: в нем теперь оказывается воды $9\frac{1}{11}$, а вина $\frac{10}{11}$ ложки; во втором стакане останется $\frac{10}{11}$ ложки воды и $9\frac{1}{11}$ ложки вина. Итак, мы видим, что количество вина в первом стакане в точности равно количеству воды во втором стакане.

---

[1] «Будущее математики».

Этот результат может показаться случайным, но если вы измените условия задачи, вы убедитесь, что всегда будут получаться одинаковые количества. Если дано x ложек воды и x ложек вина, то количество вина, которое попадет в первый стакан, неизменно будет оказываться равным количеству воды, попавшему во второй. Даже если начать с неравных количеств жидкости в стаканах, например с x ложек воды и y ложек вина, и затем проделать все процедуры, как в первоначальной задаче, то в конце концов вина в первом стакане будет столько же, сколько воды во втором.

Это хороший пример математической безвкусицы. При четком и хорошем математическом доказательстве результат не появляется неожиданно в последней строчке, его логическая неизбежность должна быть видна на протяжении всего хода доказательства.

Приведенная выше тривиальная задача использует своего рода маскировку. *В ней даются ненужные сведения,* отвлекающие ваше внимание от существа задачи. Лишним условием является предложение: «Смесь энергично взбалтывается». На самом деле существенно лишь то, что мы переливаем ложку жидкости из первого стакана во второй, а затем ложку смеси из второго стакана в первый. Совершенно не важно, какую жидкость переливают, важно лишь, что в конце концов оба стакана содержат такое же количество жидкости, как и в самом начале. А если это действительно так, то в первый стакан влили ровно столько вина, сколько нужно, чтобы заместить взятую оттуда воду. Конечно же, вода, взятая из первого стакана, обнаружится во втором. Количества жидкости должны быть равны; и ни дроби, ни алгебра здесь ни при чем.

В общем виде эта задача выглядит так: имеется стакан воды и стакан вина, мы производим ряд операций с этими жидкостями так, что в итоге количество жидкости в каждом стакане равно первоначальному, тогда количество воды, попавшее в вино, должно равняться количеству вина в воде.

Это настолько очевидно, что едва ли стоит об этом говорить. Общий вид задачи гораздо проще, чем проделанные ранее вычисления. Область применения задачи в общем виде гораздо шире. Вы можете переливать ложки жидкости из одного стакана в другой и обратно сколько угодно раз, и все же общий принцип останется неизменным.

Таким образом, исследование задачи состоит в том, чтобы отбросить все лишние данные и оставить только существенно необходимые условия. Чем меньше данных, тем легче найти решение. Общая теорема редко содержит что-нибудь запутанное; ее цель — *обратить ваше внимание на действительно важные факты.*

В элементарной математике мы встречаем смесь всяких важных и неважных деталей. В высшей математике мы пытаемся разделить различные элементы и изучить каждый в отдельности. В этом смысле высшая математика, пожалуй, гораздо проще, чем элементарная.

По всей вероятности, самый известный пример упрощения через обобщение представляет теорема Гильберта о конечном базисе. В 1868 году Гордан путем трудоемких вычислений доказал некую теорему, которую я здесь не собираюсь излагать. Она утверждала, что некоторые специальные совокупности многочленов обладают определенным свойством. В 1890 году Гильберт доказал этот же результат очень просто и без всяких вычислений. Это ему удалось сделать потому, что он отбросил 90 процентов информации, использованной Горданом. Он доказал также, что теорема справедлива не только для этих специальных классов многочленов, но и для любых классов многочленов вообще!

Мы перешли от забавного к великому. Насколько тривиальна и ограничена задачка о воде и вине, настолько глубока и всеобъемлюща теорема Гильберта. Оба эти случая иллюстрируют один и тот же принцип: «Бóльшая степень обобщения и бóльшая простота неотделимы друг от друга» — и тем самым еще раз показывают огромные возможности математики сводить под одну крышу совершенно разные объекты.

Но нельзя судить о важности какого-либо математического исследования по отдельным частным задачам. Примером тому может служить топология. Топологию иногда называют «резиновой геометрией» — геометрией фигур, нарисованных на растягивающемся листе. В некотором смысле это так — она действительно рассматривает свойства таких фигур. Но ее значение связано с тем, что на эластичном листе нет постоянной длины. Там не может быть теорем, подобных теореме Пифагора. Можно лишь делать выводы типа: «эта кривая непрерывна, а та — разделена на две отдельные части». Непрерывность — это основное свойство, изучаемое топологией. В ней говорится о таких свойствах, которые сохраняются при *непрерывных преобразованиях*. Так как существует очень много вещей, допускающих такого рода преобразования, топология охватывает широкий круг вопросов. Она все больше привлекает как чистых математиков, так и инженеров; с ее помощью можно доказать ряд удивительных результатов. Она сочетает в себе наивысшую степень обобщения с максимальной простотой.

УНИФИКАЦИЯ

Все, о чем мы говорили выше, имело целью расширить область вопросов, подвластных математике. Исследование, открытие закономерностей, объяснение смысла каждой закономерности, изобретение новых закономерностей по образу уже известных — все эти виды деятельности расширяют область действия математики. С практической точки зрения становится исключительно трудным следить за всеми полученными результатами, и нельзя сказать, чтобы нагромождение не связанных между собой теорем представляло отрадное зрелище. Будучи и деловыми людьми и художниками одновременно, математики чувствуют потребность собрать все эти разрозненные результаты.

Не удивительно, что вся история математики состоит из чередующихся процессов «расширений» и «сокращений». Например, внимание математиков привлекает какая-нибудь задача, пишутся сотни статей, каждая из которых освещает лишь одну сторону истины. Вопрос разрастается. Затем какой-нибудь гений, опираясь на все данные, собранные с таким трудом, заявляет: «Все, что мы знаем, станет почти очевидным, если посмотреть на это вот с такой точки зрения». После этого никому, кроме историков математики, нет уже необходимости изучать сотни отдельных статей. Разрозненные выводы объединяются в одну простую доктрину, важные факты отделяются от шелухи, и прямой путь к желаемому выводу открыт для всех. Объем сведений, которые нужно изучать, сократился. Но это еще не конец. После того как новый метод стал всеобщим достоянием, возникают новые вопросы, для решения которых он недостаточен, и снова начинаются поиски ответов, снова публикуются статьи, снова начинается процесс «расширения».

Если бы можно было свести все знания к двум общим законам, математик не был бы удовлетворен. Он не успокоился бы до тех пор, пока не доказал бы, что оба эти закона основываются на одном принципе. Но тогда он стал бы несчастным, так как ему нечего было бы делать. К счастью, перспектива такого застоя совершенно невероятна; решение одной проблемы всегда создает новую проблему. Всегда есть и всегда будет материал для изучения и трудности, которые нужно преодолевать.

Каким путем это происходит, можно видеть на замечательном примере унификации, относящемся примерно к 1900 году. Тогда обнаружилось, что большинство ранее изучавшихся функций являются частными случаями одного класса функций — класса гипергеометрических функций. С тех пор теория гипергеометрических функций является мощным средством объединения разрозненных математических знаний. Затем эти функции были изучены с новых сторон. Было, между прочим, доказано, что их особые свойства обусловлены тем, что они имеют три особые точки (незачем объяснять здесь, что такое «особая точка»). Итак, функции с простыми свойствами имеют три или менее особые точки; именно такими функциями занималась классическая математика. Но здесь сразу возникал вопрос: какими же свойствами будет обладать функция с *четырьмя* особыми точками? Этот вопрос по существу не решен до сих пор.

И так будет всегда. И если бы оказалось, что вся существующая математика рассматривает только явления со свойствами *A*, *B* и *C*, математики немедленно спросили бы: «А что случится, если какой-нибудь предмет будет обладать только одним из этих свойств? А если ни одним из них?» И снова погрузились бы в работу.

Глава третья

ЗАКОНОМЕРНОСТИ В ЭЛЕМЕНТАРНОЙ МАТЕМАТИКЕ

Наше благополучие — лишь равновесие обстоятельств, и мудрость состоит в том, чтобы не дать непредвиденному опрокинуть это равновесие.

Р. Бриджес, Завещание красоты

Давайте на минуту посмотрим на различные закономерности с утилитарной точки зрения, а именно, с позиции человека, которому просто нужно сдать экзамен. Самый важный вопрос для экзаменующегося — это характер экзаменатора: что его интересует? Какие качества экзаменуемого он пытается проверить?

Многих экзаменаторов интересует только способность учащихся производить шаблонные операции: знает ли он таблицу умножения? Умеет ли пользоваться логарифмами? Умеет ли применять сотни других обычных приемов? Несомненно, что проверять знание учащимися обычных операций необходимо. Но еще с детства я считаю подобные проверки самыми скучными, самыми тупыми экзаменами. И однако их очень любят посредственные учителя, которые считают, что их задача сводится просто к тому, чтобы натаскать класс на упражнениях «от сих до сих».

Другие экзаменаторы стараются поощрять более инициативных учеников, пытающихся воспринять не только факты, но и вкус к предмету. Эти экзаменаторы хотят проверить предприимчивость, воображение и инициативу учащихся, поэтому они выискивают задачи, с помощью которых можно было бы проверить эти качества.

Некоторые учителя считают, что это несправедливо, они считают, что учеников нельзя подготовить к подобным непредсказуемым экзаменам. Но это неверно. Конечно, нельзя, например, предсказать ход сражения. Но разве поэтому невозможно готовить военные кадры? Обучение офицеров лишь частично основывается (или должно основываться) на общих принципах, характерных для всех сражений; другой частью должно служить развитие их собственной инициативы. С этой целью будущий офицер на занятиях оказывается в целом ряде сложных непредвиденных ситуаций, где он должен проявить находчивость и инициативу. Точно такой же подход возможен и к мирному обучению математиков. Есть много общего между экзаменом и сражением.

Экзаменационный билет может содержать как рядовую задачу, так и проблемный вопрос, требующий сообразительности. Откуда знать студенту, какого рода задача перед ним? Ведь задачи не всегда различаются по внешнему виду.

Возьмем, например, следующую систему уравнений:

$$\left.\begin{aligned} 127x + 341y &= 274 \\ 218x + 73y &= 111 \end{aligned}\right\}.$$

Имеется шаблонный путь решения таких систем. Если мы умножим первое уравнение на 73, а второе на 341 и затем вычтем первое из второго, то получим $x = \frac{17\,849}{65\,067}$. Это выражение сразу не упрощается, поэтому нельзя обойтись без арифметических вычислений. Экзаменатор, который задает подобный вопрос, просто хочет проверить, знает ли учащийся обычный способ решения системы уравнений с двумя неизвестными и хватит ли у него терпения и внимания, чтобы правильно проделать все арифметические вычисления. (Подразумевается, что требуется точный ответ в виде дроби. Если же достаточно получить приближенный результат, то, разумеется, можно сократить все расчеты, воспользовавшись логарифмической линейкой.) Значение y вычисляется аналогично.

Для контраста рассмотрим другую систему уравнений:

$$\left.\begin{aligned} 6751x + 3249y &= 26\,751 \\ 3249x + 6751y &= 23\,249 \end{aligned}\right\}.$$

Здесь числа более громоздкие, но задача гораздо легче. Она не имеет целью проверить знание обычных методов решения. Ее можно решить очень просто, если, конечно, экзаменующийся поймет, как это сделать.

Что же именно подсказывает нам, что этот пример требует сообразительности? Во-первых, здесь примечателен выбор чисел, но особенно бросается в глаза закономерность в левых частях обоих уравнений:

$$\begin{matrix} * & \bullet \\ \bullet & * \end{matrix}\;.$$

Левые части уравнений можно представить в алгебраическом виде:

$$ax + by, \tag{1}$$

$$bx + ay. \tag{2}$$

Когда я рассматриваю эти выражения, у меня возникает довольно-таки странный ход мыслей: а что, если переименовать наши неизвестные? Там, где стоял x, поставим y, и наоборот; тогда первое уравнение станет вторым, а второе примет вид первого. Значит, выражения (1) и (2) равноценны. Поэтому было бы неправильным предпринимать что-либо с выражением (1) и не делать этого же с (2). Значит, какие бы операции мы ни проводили, они должны быть проделаны с обоими выражениями.

Какого же рода операции это могут быть? Очевидный шаг — сложение выражений (1) и (2). Это совершенно симметричная операция.

Какие еще симметричные операции можно проделать? Можно было бы, конечно, перемножить (1) и (2), но это нас нисколько не приблизит к решению. Нам нужна операция типа $m(1)+n(2)$. На первый взгляд кажется, что надо взять $m=n$. В противном случае одно из выражений получит больший множитель, а второе — меньший. Однако имеется и другой путь.

Уравнение $p=q$, конечно, симметрично по отношению к p и q. Но если мы перепишем его так, чтобы все ненулевые члены находились в левой части, то мы получим:

$$p-q=0.$$

Здесь нет симметрии. Мы произвольно взяли p со знаком плюс и q со знаком минус. Но если бы мы взяли наоборот, то получили бы $q-p=0$, что тоже выражает $p=q$. Следовательно, хотя выражение $p-q$ несимметрично, уравнение $p-q=0$, являющееся другой записью уравнения $p=q$, следует рассматривать как симметричное. И мы не погрешим против симметрии, если, имея два равноправных выражения, вычтем одно из другого.

Вернемся к исходному примеру. Теперь у нас есть две операции, обладающие симметрией, подходящей для нашей задачи,— сложение и вычитание данных уравнений. Произведя эти операции, мы получим:

$$\left.\begin{aligned} 10\,000x + 10\,000y &= 50\,000 \\ 2502x - 2502y &= 2502 \end{aligned}\right\}$$

или

$$\left.\begin{aligned} x+y&=5 \\ x-y&=1 \end{aligned}\right\},$$

откуда

$$\left.\begin{aligned} x&=3 \\ y&=2 \end{aligned}\right\}.$$

Учащиеся могут получить истинное удовольствие при решении подобного примера. Конечно, не все пойдут тем путем, который я описал. Но всякий, кто будет этот пример решать без вычислений, непременно почувствует закономерность этой системы уравнений. Победа над задачей зависит от чувствительности к закономерностям. Такое сочетание военного и художественного подхода нигде, пожалуй, не осуществляется в такой степени, как в математике.

В задачах, которые ставит перед нами жизнь, экзаменатором является сама природа. И здесь опять чрезвычайно важно уметь определить, требует ли данная частная задача шаблонного подхода или она обладает каким-нибудь особым свойством, которое дает возможность найти более простое решение. Конечно, много практических задач можно решить только обычными методами, так как они лишены закономерностей; в частности, если в них присутствует элемент случайности, беспорядочности. В топографии, например, при определении расположения городов принимается во внимание масса геологических и исторических причин. Трудно ожидать каких-либо изящных соотношений между расстояниями на карте — здесь вычисления неизбежны. С другой стороны, в фундаментальной научной проблеме (относящейся к строению атома или к Вселенной) мы надеемся (имея или не имея на это основание) найти скрытую простоту; основные физические теории обычно обладают математическим изяществом, хотя их практическое применение к сложным проблемам, что и говорить, не всегда этим отличается.

ВОССТАНАВЛИВАЕМ ХОД МЫСЛИ ЭКЗАМЕНАТОРА

Бывает, что палеонтологи откапывают окаменелую косточку и по ней начинают восстанавливать очертания какого-нибудь ископаемого животного. Подобным образом можно представить себе деятельность экзаменатора, причем экзаменационные вопросы играют здесь роль окаменелой кости.

Хороший экзаменационный вопрос — не такое уж простое дело; он должен содержать интересный замысел или приводить к неожиданному ответу. Составлять такие вопросы нелегко. Поэтому экзаменатор, занимающийся исследовательской работой, старается обычно подметить материал, который он мог бы использовать в качестве экзаменационного вопроса. Часто в довольно серьезной работе встречаются алгебраические выкладки, которые можно вырвать из контекста и сформулировать как самостоятельную задачу.

Например, несколько лет назад студенты обратились ко мне со следующим примером из экзаменационного билета, который они затруднялись решить.

«Докажите, что если

$$\frac{ac-b^2}{a-2b+c}=\frac{bd-c^2}{b-2c+d},$$

то обе эти дроби равны выражению

$$\frac{ad-bc}{a-b-c+d}\text{»}.$$

Пример этот имеет очень определенную форму, и очевидно, что решение при помощи длинных и бесформенных вычислений ни на шаг не приблизило бы нас к ответу. Меня же в этом случае больше всего интересовал вопрос: каким образом экзаменатор додумался до такой задачи?

Внешний вид задачи позволяет нам увидеть следующее: $ac-b^2=0$ есть условие того, что три числа a, b, c составляют геометрическую прогрессию; выражение $ac-b^2$ содержится в числителе. Подобное же выражение стоит в числителе второй дроби. В знаменателе мы видим выражения $a-2b+c$ и $b-2c+d$, которые ассоциируются с арифметическими прогрессиями: $a-2b+c=0$ является условием того, что a, b и c — члены арифметической прогрессии. Еще здесь имеется своего рода правило, по которому знаменатели можно вывести из числителей. Например, в третьей дроби мы видим наверху произведение чисел a и d; а внизу — сумму этих чисел (числитель содержит ad, а знаменатель $a+d$). Отрицательные члены связаны подобным же образом; наверху стоит $-bc$, внизу $-(b+c)$. Это правило с таким же успехом применимо и к первым двум дробям: в первой дроби, например, мы видим в числителе $-b^2$, т. е. $-bb$, а в знаменателе $-2b$, т. е. $-(b+b)$.

Почти невозможно изобрести задачу, где элементы были бы так переплетены между собой. Такие вещи не изобретаются, с ними можно только нечаянно столкнуться. Я был совершенно уверен, что экзаменатор искал условия для чего-то и эти дроби образовались у него в процессе работы.

С чего начать решать эту задачу, было довольно ясно, нужно было ввести новый символ k для обозначения дробей. Тогда задача сформулировалась следующим образом:

$$\frac{ac-b^2}{a-2b+c}=k \tag{1}$$

и

$$\frac{bd-c^2}{b-2c+d}=k; \tag{2}$$

докажите, что

$$\frac{ad-bc}{a-b-c+d}=k. \tag{3}$$

Ввести новый символ k, когда имеешь дело с равными дробями,— дело обычное. Но вот, что делать дальше, было совсем неясно. Я перепробовал разные методы, которые хотя и вели к доказательству, но не удовлетворяли меня. Я продолжал обдумывать эту задачу на досуге и примерно неделю спустя натолкнулся на следующее. Уравнение (1) можно представить в виде $ac-k(a+c)=b^2-2bk$. Само собой напрашивается теперь ввести в обе части уравнения, чтобы придать им законченный вид, дополнительный член k^2. Справа это будет дополнение до полного квадрата $(b-k)^2$, а слева получится произведение $(a-k)\ (c-k)$. Итак, $(a-k)\ (c-k)=(b-k)^2$. Иными словами, *уравнение* (1) *выражает тот факт, что* $a-k$, $b-k$ и $c-k$ — *члены геометрической прогрессии.*

Вот теперь все ясно. Уравнение (2) показывает, что $b-k$, $c-k$ и $d-k$ являются также членами геометрической прогрессии, следовательно, $a-k$, $b-k$, $c-k$ и $d-k$ составляют геометрическую прогрессию. Но произведение первого и четвертого членов геометрической прогрессии равно произведению ее второго и третьего членов. (Пусть члены геометрической прогрессии будут A, AR, AR^2, AR^3; тогда $A \cdot AR^3 = AR \cdot AR^2$.) Итак, мы имеем:

$$(a-k)\ (d-k) = (b-k)\ (c-k).$$

Если мы откроем скобки, вычеркнем k^2 и разрешим получившееся линейное уравнение относительно k, то получим равенство (3).

Можно предположить, что в своей научной работе экзаменатор стал перед вопросом: «Каким условиям должны удовлетворять четыре числа a, b, c и d, чтобы разности между каждым из них и некоторым постоянным числом k составляли геометрическую прогрессию?» Этот вопрос он и задал на экзамене.

Разобранный выше пример поучителен не только с точки зрения экзаменационных вопросов. Он также подтверждает тезис: *«Там, где есть закономерность, есть и смысл»*. Если в математической работе какого-либо рода повторяется некоторая ярко выраженная закономерность, всегда необходимо исследовать, *почему* она встречается. Она обязательно имеет какое-нибудь значение, которое мы можем воспринять как идею, а не просто как набор символов. Открыть теорему и доказать ее путем бесформенных вычислений — этого совершенно недостаточно. Это означает, что мы не понимаем того, что открыли.

Иногда требуется много времени, чтобы понять смысл алгебраической формулы; обычно имеется так много возможных методов подхода, что неизвестно, какой из них правильный. Я нахожу, что, имея дело с такими задачами, мой мозг совершает замедленную работу: сначала я не знаю, как подступиться к такой задаче, а через день, неделю, а иногда и месяц приходит вдохновение. Если учащийся должен решить подобную задачу в течение трех часов, т. е. в условиях экзамена, многое еще зависит от удачи. Обычный метод преодоления подобных трудностей заключается в том, чтобы припомнить соответствующие разделы учебника или какую-нибудь более простую задачу, чтобы заставить мозг работать в нужном направлении.

ВОССТАНАВЛИВАЕМ ХОД МЫСЛИ АВТОРА

Собрать достаточное количество задач для составления учебника гораздо труднее, чем подобрать задачи для экзаменационных билетов. Лучший способ подобрать интересные вопросы по алгебре, на мой взгляд,— это изучить большое количество научных статей за последние несколько сот лет и выбрать все алгебраические результаты, которые были доказаны случайно в процессе более серьезной работы. Вероятно, самые интересные задачи в учебниках алгебры появились именно таким образом.

Просматривая любой учебник, вы заметите, что в ряде примеров отсутствует закономерность. Это примеры такого рода, которые каждый может легко составить; они относятся к такому же типу, как следующий пример из арифметики: найти сумму $27+46+39$; или из алгебры: умножьте $(5x^2-3x+7)$ на $(4x+11)$.

Конечно, такие упражнения тоже необходимы. Но есть и примеры другого типа, которые не могли быть выдуманы экспромтом, и всегда интересно попытаться угадать, как автору пришла в голову подобная задача.

Например, в одной книге по алгебре встречается следующая задача: «Докажите, что если

$$a = zb + yc,\ b = xc + za,\ c = ya + xb,$$

то

$$\frac{a^2}{1-x^2} = \frac{b^2}{1-y^2} = \frac{c^2}{1-z^2}\text{».}$$

Задача нетрудная, но в данный момент нас не интересует, как ее решить, мы хотим знать, как она была составлена. Если не предполагать небывалого совпадения, то задача возникла следующим образом: авторы [1] беседовали, но не об алгебре, а о тригонометрии. В связи с тригонометрией возникает часто целый ряд алгебраических задач. В алгебре, как известно, обычно нельзя однозначно найти три неизвестных, имея только два уравнения. Если у вас три неизвестных и три уравнения, вы можете (как правило) найти неизвестные. Если же у вас три неизвестных и четыре уравнения, то, вообще говоря, задача неразрешима. Например, можно взять следующие уравнения:

$$x=1,\ y=2,\ z=3,\ x+y+z=0,$$

противоречащие друг другу. Четыре уравнения с тремя неизвестными решаются только в том случае, если четвертое уравнение является простым повторением того, что мы сами можем узнать из первых трех, как, например,

$$x=1,\ y=2,\ z=3,\ x+y+z=6.$$

Тогда никаких затруднений не возникает.

Теперь взгляните на задачу о треугольнике (рис. 4). Даны три стороны a, b, c. Необходимо найти углы треугольника A, B, C. Мы имеем три неизвестных, значит, хватило бы трех уравнений, остальные бы только мешали. Тригонометрия в этом случае дает нам восемь уравнений.

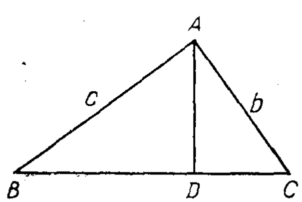


Рис. 4.

[1] Речь идет о книге «Высшая алгебра» двух авторов: Холла (Hall) и Найта (Knight).— *Прим. ред.*

Прежде всего, если мы опустим перпендикуляр AD из вершины A на сторону BC, получим $BC = BD + DC$; отсюда

$$a = c\cos B + b\cos C. \tag{1}$$

Опуская перпендикуляры из точек B и C соответственно на стороны AC и BA, получим еще два уравнения, которые я назову уравнениями (2) и (3), но приводить здесь не буду.

Трех уравнений совершенно достаточно, чтобы найти три неизвестных. С точки зрения алгебры мы имеем все, что нужно. Но тригонометрия продолжает преподносить нам свои дары. Учебник обращает наше внимание на теорему косинусов:

$$\cos A = \frac{b^2 + c^2 - a^2}{2bc} \tag{4}$$

и на аналогичные выражения для $\cos B$ и $\cos C$ (уравнения (5) и (6)). И наконец, имеется теорема синусов:

$$\frac{a}{\sin A} = \frac{b}{\sin B} = \frac{c}{\sin C};$$

здесь имеются еще два знака равенства, что дает нам уравнения (7) и (8).

Далее, мы знаем, что можно найти углы треугольника и что все получаемые ответы не противоречат друг другу. Это значит, что уравнения (4)—(8) не содержат никаких новых сведений, а лишь являются алгебраическими следствиями уравнений (1), (2) и (3).

Довольно очевидно, что уравнения (4), (5) и (6) получаются из уравнений (1), (2) и (3), когда мы разрешаем последние относительно $\cos A$, $\cos B$ и $\cos C$. Не так очевидно, что из этих уравнений следует правило синусов. Итак, перед нами не слишком очевидный результат, который требует алгебраических преобразований. Чтобы избавиться от тригонометрических функций, положим $\cos A = x$, $\cos B = y$, $\cos C = z$. Поскольку $\sin^2 A + \cos^2 A = 1$, нам придется записать, что $\sin^2 A = 1 - x^2$. Мы возводим в квадрат оба уравнения, составленные по теореме синусов, и, таким образом, полностью переходим на язык алгебры. И вот мы пришли к задаче, сформулированной в учебнике по алгебре.

В тригонометрии различные результаты с косинусами доказываются с помощью одних геометрических построений, с синусами — с помощью других. Тот факт, что теорема синусов является алгебраическим следствием теоремы косинусов, не подчеркивается.

Собственно, и теперь еще мы не определили алгебраического пути сведéния одной теоремы к другой. Но мы увидели, что *должен быть какой-то способ* доказать этот результат алгебраически. Если его нет, тригонометрия противоречит сама себе.

ПЕРЕВОД ИЗ ОДНОЙ ОБЛАСТИ В ДРУГУЮ

То, что было проделано в предыдущем параграфе, по существу является одним из видов перевода. Мы начали с известного в тригонометрии факта, что теорема синусов не противоречит теореме косинусов. Мы переводим этот факт на язык алгебры и обнаруживаем, что одна система уравнений может быть выведена из другой. От знакомого факта (во всяком случае, знакомого для любого, кто изучал школьный курс тригонометрии в течение последних двух-трех лет) мы перешли к незнакомому. Мы не просто получили новый результат, мы уточнили уже известную формулировку: вместо «не противоречит» мы теперь говорим более определенно «может быть выведена».

Перевод — очень ценное упражнение, потому что, прежде чем выразить что-либо на другом языке, вы должны уяснить себе, о чем идет речь. В различных местах этой книги мы попытаемся излагать многое так ясно, чтобы это можно было объяснить «ангелу по телефону».

Мы, например, попробуем объяснить геометрию без чертежей, без ссылок на физические представления. В самом деле, геометрию можно объяснить на языке чисел. Вам это может показаться странным, если вы привыкли думать, что геометрия изучает формы предметов. Но математика имеет дело с закономерностями и не интересуется, на каком конкретном материале они осуществляются. Наша способность объяснить геометрию «ангелу по телефону» означает просто, что мы можем, пользуясь только числами, выразить те же самые *закономерности*, что встречаются в геометрии. Подобно комментатору по радио, мы могли бы информировать «ангела» о ходе футбольного матча, сообщив ему счет 7 : 7. Однако мы могли бы передать и ощущения игрока, которого ударили по ноге.

В качестве примера перевода я попытаюсь перевести на язык алгебры тот факт, что $\cos^2\theta$ никогда не превосходит 1.

Воспользуемся методом координат. Пусть O — начало координат, P — точка с координатами (a, b), Q — точка с координатами (x, y) (рис. 5). Эти три точки образуют треугольник OPQ, и все, что можно сказать относительно этого треугольника, мы скажем, используя только числа a, b, x, y, т. е. в терминах алгебры. Длину OP удобно обозначить через p, OQ через q и PQ через r. Разумеется, отрезки p, q и r можно выразить через a, b, x и y; применяя теорему Пифагора к треугольникам ONP, OSQ и PHQ, мы получим:

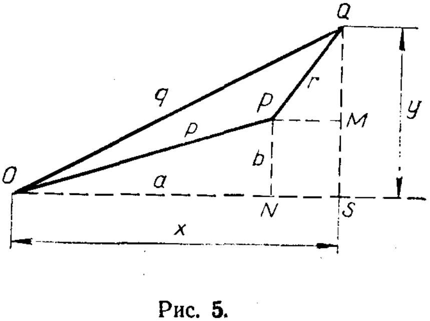


Рис. 5.

$$p^2 = a^2 + b^2,$$
$$q^2 = x^2 + y^2,$$
$$r^2 = (x-a)^2 + (y-b)^2.$$

Если угол между OP и OQ назвать θ, то $\cos\theta$ можно найти из формулы $2\,pq\,\cos\theta = p^2 + q^2 - r^2$. Подставляя сюда приведенные выше значения p, q и r, увидим, что взаимно уничтожаются все члены второй степени, за исключением $2ax+2by$. Отсюда имеем:

$$pq\cos\theta = ax+by.$$

Если мы выразим p и q через a, b, x, y, то появятся квадратные корни; поскольку, однако, мы хотим выразить тот факт, что $\cos^2\theta$ никогда не превосходит 1, нам все равно придется возводить в квадрат. Мы найдем, таким образом, что $(ax+by)^2$ никогда не превосходит p^2q^2 и, значит, $(ax+by)^2$ никогда не превосходит $(a^2+b^2)\,(x^2+y^2)$.

Это чисто алгебраический результат; мы можем выразить его смысл, сказав, что для действительных чисел разность $(a^2+b^2)\,(x^2+y^2)-(ax+by)^2$ никогда не отрицательна.

Почему же это выражение не может быть отрицательным? Данное утверждение приблизительно соответствует известному из тригонометрии предложению о том, что $1-\cos^2\theta$ не может быть отрицательным (для действительных углов θ). Но $1-\cos^2\theta=\sin^2\theta$. Теперь мы уже ожидаем, что приведенное выше алгебраическое выражение окажется квадратом. Так оно и есть. Если раскрыть это выражение, то из семи полученных членов четыре взаимно уничтожатся. Останется $(bx-ay)^2$. Это выражение не может быть отрицательным. Вот мы и получили алгебраический ответ, не прибегая к тригонометрии.

Наш результат можно представить в виде тождества

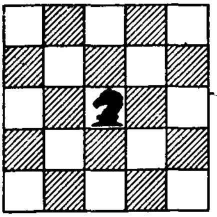
$$(a^2+b^2)\,(x^2+y^2) = (ax-by)^2 + (bx-ay)^2,$$

справедливого для всякой четверки чисел. В главе 4 мы вернемся к нему и увидим, в каком направлении можно его обобщить.

ХОД ШАХМАТНОГО КОНЯ

Рассмотрим еще один пример, иллюстрирующий одно из очень важных применений перевода, а именно представление задачи в такой форме, когда ответ можно видеть с первого взгляда. Такой перевод не меняет сущности задачи; если подходить с чисто математической точки зрения, можно сказать, что он вообще ничего не меняет. Но для нас перевод этот очень ценен, потому, что он превращает незнакомое в знакомое.

Примером такого перевода является графическое представление функции. График легче понять с первого взгляда. Здесь мы производим перевод из алгебры в геометрию. Очень часто приходится делать обратный перевод — из геометрии в алгебру; ряд примеров



1	2	3	4	5
6	7	8	9	10
11	12	13	14	15
16	17	18	19	20
21	22	23	24	25

Рис. 6.

такого типа вы найдете в этой книге. Можно также переводить одну геометрическую задачу в другую, но так, чтобы неизвестный тип задачи был переведен в известный.

Рассмотрим задачу с шахматным конем. Конь стоит в центре квадратной доски из 25 полей (рис. 6). Конь должен двигаться так, чтобы побывать в каждой клетке один и только один раз.

Размышляя над этой задачей, я нахожу довольно трудным отчетливо представить себе в уме, попадает ли наш конь в затруднительное положение, скажем, после дюжины ходов; глядя на клетки, на которых конь еще не был, я не могу сказать, соединятся ли они в простую цепь ходов. Нельзя ли несколько преобразовать нашу задачу, с тем, чтобы было яснее, что мы делаем? С точки зрения коня клетки 2 и 13 являются соседними: он может попасть из клетки 13 в клетку 2 за один ход; но, скажем, клетки 12 и 13 соседними не являются, так как конь не может перейти из одной в другую за один ход. Поэтому, если нас интересует только задача о ходе конем, мы можем совершенно забыть о действительной форме шахматной доски; для решения этой задачи достаточно нарисовать диаграмму из 25 клеточек и разместить их таким образом, чтобы поле 2 находилось рядом с полем 13, а поле 12 дальше от него. (В самом деле, конь должен сделать три хода, прежде чем он попадает из 13 в 12).

Получившаяся диаграмма представлена на рис. 7. Числа, соответствующие клеткам, переход между которыми совершается за один ход, соединены отрезками. Точки пересечения этих отрезков не имеют значения. На рисунке, сделанном на листе бумаги, невозможно избежать таких случайных пересечений. Их можно было бы избежать в пространственной (трехмерной) модели, соединив клетки мягкими проводами; эти провода мы могли бы изогнуть так, чтобы они не касались друг друга.

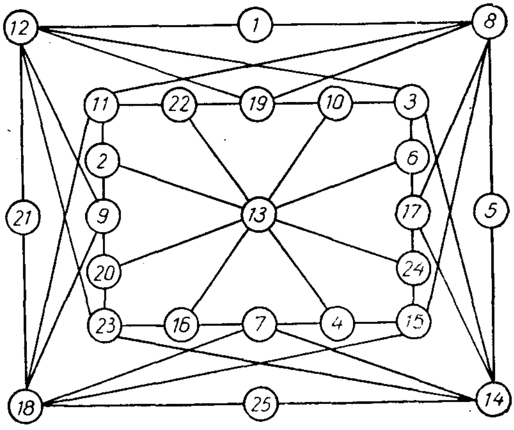


Рис. 7.

Нарисовать такую диаграмму не такое уж простое дело, как это кажется на первый взгляд. Нужно проявить достаточно внимательности, чтобы не получилась слишком запутанная сеть линий. Главный принцип, которого я придерживался,— это соблюдение симметрии. Клетка 13 должна быть помещена в центре, а остальные — вокруг нее, но так, чтобы сохранить симметрию шахматной доски.

Теперь очень легко выбрать маршрут коня. Он может, например, отправиться из 13 в 10, затем пройти по «внутреннему кругу» (19, 22, 11, 2, 9, 20, 23, 16, 7, 4, 15, 24, 17, 6, 3) и перейти на «внешний круг» (12, 21, 18, 25, 14, 5, 6, 1).

Нам не нужно вдаваться в теорию ходов шахматного коня. Пример выбран не потому, что данная задача важна, а затем, чтобы показать, как умение наглядно представить себе суть задачи упрощает отыскание решения.

УНИКАЛЬНАЯ ЗАКОНОМЕРНОСТЬ

В главе десятой, посвященной проективной геометрии, мы встретимся с функцией $\frac{(a-b)(c-d)}{(b-c)(d-a)}$. Это так называемое сложное отношение четырех чисел a, b, c, d. Поскольку это довольно длинное выражение, будем в дальнейшем обозначать его через $f(a, b, c, d)$, что несколько короче.

Порядок букв a, b, c, d важен. Например, если мы записали бы $f(a, d, c, b)$, это означало бы $\frac{(a-d)(c-b)}{(d-c)(b-a)}$, что вовсе не равно $f(a, b, c, d)$. В самом деле, нетрудно видеть, что выражения $f(a, b, c, d)$ и $f(a, d, c, b)$ взаимно обратны. Если $f(a, b, c, d) = x$, то $f(a, d, c, b) = \frac{1}{x}$.

Имеются 24 способа, которыми могут быть записаны подряд 4 буквы, следовательно, мы можем получить 24 различных выражения, записывая буквы в различном порядке; например, $f(a, c, d, b)$, $f(d, c, b, a)$, $f(d, b, a, c)$ и т. д. Если у вас хватит терпения написать все 24 выражения, вы увидите, что в действительности не все они различны. Всего встречается 6 различных значений, каждое из них по 4 раза. Вы можете проверить это алгебраически или, если угодно, можете даже вычислить значения выражений, положив, например, $a = 0$, $b = 1$, $c = 2$, $d = 3$. Размещая величины 0, 1, 2, 3 в различном порядке, вы получите значения $-\frac{1}{3}$, -3, $\frac{1}{4}$, 4, $\frac{3}{4}$, $\frac{4}{3}$, которые может принимать функция; каждое значение может быть получено четырьмя различными способами, например: $-\frac{1}{3} = f(0, 1, 2, 3) = f(1, 0, 3, 2) = f(2, 3, 0, 1) = f(3, 2, 1, 0)$.

Шесть значений, найденных таким образом, связаны между собой. Если мы возьмем любое из них и найдем для него обратную величину, то получим другое значение из нашей группы; если мы вычтем какое-либо значение из 1, то получим еще одно значение, принадлежащее к нашей группе, Например, если мы начинаем с $-\frac{1}{3}$, то, обращая, получаем -3. Вычитая -3 из 1, получаем 4. Находим обратную величину, получаем $\frac{1}{4}$, вычитаем $\frac{1}{4}$ из 1, получаем $\frac{3}{4}$, и, наконец, обращая последнее выражение, получаем $\frac{4}{3}$.

Таким образом, имея только одно значение, мы можем найти все остальные. Если мы начинаем с x, то, обращая, получаем $\frac{1}{x}$. Вычитая из 1, получаем $1-\frac{1}{x}$, что можно записать в виде $\frac{x-1}{x}$. Обращая, получаем $\frac{x}{x-1}$, вычитая из 1, получаем $\frac{1}{1-x}$, и, наконец, беря обратную величину, получаем $1-x$.

Вы, возможно, думаете, что, произведя подобные операции еще несколько раз, мы получим какие-нибудь новые выражения. Но это не так. Сколько бы раз вы ни производили указанные операции, вы не выйдете из замкнутого круга значений x, $\frac{1}{x}$, $1-x$, $\frac{1}{1-x}$, $\frac{x-1}{x}$ и $\frac{x}{x-1}$.

Если теперь вам скажут, что $f(a, b, c, d)$ равно, допустим, 5, вы уже будете знать, что $f(a, d, c, b)$ равно $\frac{1}{5}$, $f(a, c, b, d)$ равно -4 (потому что $1-5=-4$), *хотя вы и не знаете, чему равны a, b, c и d.* Вполне естественно, что одного только уравнения $f(a, b, c, d)=5$ недостаточно для определения значений четырех величин a, b, c и d.

Рассмотренная функция замечательна своей уникальностью. *Ни одна другая функция не обладает подобными свойствами.* Здесь мы, конечно, не принимаем во внимание простые видоизменения самой функции. Например, функция $3f(a, b, c, d)-2$ обладает теми же самыми свойствами. Но эту функцию мы рассматриваем скорее как видоизмененную старую функцию, чем как совершенно новую.

Далее, существуют сравнительно тривиальные пути построения функций, обладающих сходными свойствами. Рассмотрим симметричную функцию, скажем $\varphi(a, b, c)=a+b+c$. Сумма трех чисел не зависит от порядка написания чисел. Поэтому, если нам известно, что $\varphi(a, b, c)=5$, то мы знаем, что и $\varphi(a, c, b)$ тоже равно 5, и в каком бы порядке мы ни написали буквы a, b и c, их сумма всегда будет равна 5. Но это слишком очевидно, чтобы быть интересным.

Несколько менее очевидными являются функции такого типа, как

$$F(a, b, c)=(a-b)(a-c)(b-c).$$

Если поменять местами любые две буквы, функция изменит знак.

Так, $F(a, b, c) = -F(a, c, b) = -F(c, b, a) = -F(b, a, c) = F(b, c, a) = F(c, a, b)$. Эта функция принимает только два значения: x и $-x$.

Сколько букв мы бы ни взяли, мы всегда можем составить симметричные функции типа φ, которые остаются неизменными при любой перестановке букв, и функции типа F, которые могут принимать только два значения при перестановке букв.

Но если нам нужно, чтобы функция принимала более чем два значения, и нам хочется предсказать все остальные значения, когда мы знаем одно из них, функция должна содержать четыре буквы и должна либо совпадать со сложным отношением $f(a, b, c, d)$, либо быть вариантом этой функции (вроде рассмотренного выше $3f(a, b, c, d)-2$). В этом смысле число 4 и сложное отношение $f(a, b, c, d)$ выделяются из всех остальных чисел и функций.

Интересно, что уникальность функции $f(a, b, c, d)$ может быть доказана *без каких бы то ни было вычислений* путем одних лишь логических умозаключений. Доказательство очень короткое и основывается на теории групп.

Уникальные функции всегда стоит изучать. Если путь в какой-то город ведет через один определенный мост, то движение на этом мосту будет очень оживленным. Точно так же, если целый ряд свойств присущ только какой-то одной функции, то любая задача, касающаяся этих свойств, должна решаться с помощью этой функции. Сложное отношение $f(a, b, c, d)$ встречается во многих областях математики.

Вам может показаться интересным проверить, если вы еще этого не сделали, с помощью элементарной алгебры свойство функции $f(a, b, c, d)$, о котором говорилось в этом разделе: если $f(a, b, c, d) = x$, то все остальные 23 размещения четырех элементов a, b, c, d дают значения, являющиеся указанными выше функциями от x. На это потребуется много времени и труда. Ниже, в главе 10, когда мы узнаем геометрический смысл сложного отношения, мы сможем доказать этот результат с гораздо меньшим трудом. Тем более вы оцените ту помощь, которую геометрия оказывает алгебре, если вы произведете хотя бы некоторые из этих проверок более элементарным методом.

Между прочим, человеку, изучающему алгебру, часто полезнее решить одну и ту же задачу тремя различными способами, чем решить три-четыре различные задачи. Решая одну задачу различными методами, можно путем сравнений выяснить, какой из них короче и эффективнее. Так вырабатывается опыт.

Глава четвертая

ОБОБЩЕНИЕ В ЭЛЕМЕНТАРНОЙ МАТЕМАТИКЕ

Начиная с 1800 года, когда математика вступила в современную стадию, наметилась упорная тенденция к усилению роли абстракции и обобщения... Абстракция и обобщение, направленные на создание универсальных методов и законченных теорий, стали в повестку дня.

Э. Т. Белл, Развитие математики

Во второй главе мы говорили об обобщении, как об одном из важнейших процессов, ведущих к развитию математики.

Но обобщение не только расширяет сферу математики. Оно помогает выработать единую точку зрения. Часто незнакомый результат можно рассматривать как обобщение знакомого. Это помогает осознать новый вывод и связать его с уже известными выводами.

Например, в начале курса алгебры ребенок уже знает, что $x-a$ является множителем для выражения x^2-a^2. Позже он узнает, что $x-a$ является множителем для x^3-a^3 и вообще для выражения x^n-a^n при любом целом n. Этот старый, хорошо знакомый результат является как бы колышком, за который цепляются новые результаты. Он не объясняет и не доказывает новые результаты, но помогает нашему мозгу воспринять их. Очень часто главные трудности в учебе совсем не логического характера. Человек прекрасно видит каждый шаг, понимает, что доказательство логично, но у него остается упорное ощущение, что он не понимает, в чем состоит новый результат и к чему все это.

Например, мы начинаем изучать в школе тригонометрические функции. Они связаны с окружностью, к которой мы настолько привыкли, что, проделав несколько упражнений, мы чувствуем, что понимаем, что такое синус и косинус, хотя нам и приходится пользоваться таблицами, чтобы найти их значение. Позже математик сталкивается с бесселевыми функциями. Свойства бесселевых функций изучены, доказательства очень логичны, и все-таки у студента зачастую остается чувство, что он не знает, что из себя представляет бесселева функция. Он не может понять ее так, как понимает обычный синус.

Лично я чрезвычайно обрадовался, когда открыл, что бесселеву функцию можно рассматривать как обобщение синуса; я не могу ручаться, что эта мысль обрадует так же и других людей. Это относится скорее к психологии, чем к логике. Как бы то ни было, судите сами, чего эта мысль стоит.

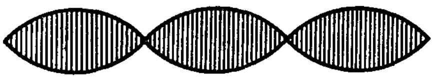


Рис. 8.

Из акустики известно, что туго натянутая проволока или струна (например, фортепьянная струна) колеблется (при определенных условиях) по синусоиде. Фортепьянные струны изготавливаются из однородной проволоки, и напряжение по всей длине струны одинаково.

Но можно ведь изготовить струну переменной толщины, так, чтобы напряжения в различных ее частях были различными. Изготовить струну переменной толщины — это вопрос технологии; а переменное напряжение можно было бы получить, например, подвесив струну вертикально или другим способом, но в это нам незачем вдаваться. Более того, можно сделать так, чтобы и толщина струны и напряжение изменялись постепенно. Мы могли бы взять обычную фортепьянную струну и, постепенно покрывая ее дополнительным материалом, сделать утолщения в некоторых местах. С помощью довольно сложных приспособлений можно было бы осуществить и постепенное изменение напряжений. Тогда колебания подобной струны постепенно переходили бы от синусоидальной формы к какой-то другой. Колебания, которые можно было бы получить подобным образом, как раз и описываются с помощью бесселевой функции. Более того, целый ряд свойств обычной синусоиды сохранился бы при таком постепенном процессе, так что многими своими свойствами, например волнообразностью графика, бесселева функция обязана синусу.

Между прочим, форма колебаний кожи барабана может быть описана с помощью бесселевых функций.

При помощи натянутой фортепьянной струны можно получить не только бесселевы функции, но и многие другие функции, с которыми приходится встречаться в математической физике. Любую из них можно рассматривать, в этом смысле, как обобщение синуса. Большую помощь в изучении этих функций оказывают постоянно обнаруживающиеся сходные свойства.

ПРИМЕР АЛГЕБРАИЧЕСКОГО ОБОБЩЕНИЯ

В алгебре есть раздел, который в одних странах проходят в школе, а в других — на первом курсе университета. Он является прекрасным примером обобщения.

Этот раздел можно было бы считать неподходящим для изложения в популярной книге по математике по ряду причин: (1) он находится почти на университетском уровне; (2) он связан или может быть связан с громоздкими алгебраическими выражениями; (3) он не очень легко дается учащимся.

Тем не менее я собираюсь его рассмотреть. На приведенные выше возражения можно ответить следующим образом: (1) из алгебры нужны лишь элементарные сведения, и нет необходимости в длинных вычислениях; (2) хотя встречающиеся алгебраические выражения и громоздки, они построены по весьма определенному принципу, который легко воспринимается; (3) в основе раздела лежит одна-единственная идея, о которой обычно не упоминают в учебниках. Эта идея заключается в следующем: надо попытаться обобщить тот факт, что две точки можно соединить одной и только одной прямой.

В такой формулировке задача, конечно, относится к геометрии, и мы можем начать с поисков геометрического обобщения. Почти все знают, что через любые три точки (не лежащие на одной прямой) можно провести окружность; люди с более широкими математическими познаниями, возможно, знают, что через четыре точки можно провести параболу, а через пять — эллипс. Но дальше мы оказываемся в тупике. Мы не можем придумать, какая же кривая проходит через шесть точек.

Попытаемся перейти на язык алгебры, как мы уже поступали в третьей главе. Это довольно легко. Известно, что любую прямую (за исключением вертикали) можно представить уравнением $y=mx+c$, подобрав соответствующим образом константы m и c. Точка определяется парой чисел. Наш геометрический результат, переведенный на язык алгебры, теперь будет звучать так: «Всегда можно подобрать два числа m и c так, чтобы прямая $y=mx+c$ проходила через две заданные точки (a, p) и (b, q)». Нужно добавить, что a и b должны быть неравны, потому что в противном случае точки лежали бы на одной вертикальной прямой и коэффициент m стал бы бесконечным.

Вместо слов «прямая проходит через точки» мы могли бы написать уравнения:

$$p = ma + c, \tag{1}$$

$$q = mb + c, \tag{2}$$

которые означают, что выражение $mx+c$ принимает значение p при $x=a$ и значение q при $x=b$.

Вот и появилась алгебраическая точка зрения на задачу. В нашем распоряжении две величины m и c и два уравнения, которым они должны удовлетворять. Решить эти уравнения не представляет никакого труда.

Можно также предположить, что если мы хотим провести кривую через три точки, нам нужно брать ее уравнение в виде $y=gx^2+hx+k$ и искать три неизвестных g, h, k. Идея, конечно, правильная, но ее надо обосновать.

Встречаются, к сожалению, и такие случаи, когда система трех уравнений с тремя неизвестными не имеет решения. Рассмотрим, например, уравнения

$$\left.\begin{aligned} g-h&=1\\ h-k&=2\\ k-g&=3 \end{aligned}\right\}.$$

У этой системы нет решения. Первое уравнение говорит нам, что g больше h, второе — что h больше k, а третье — что k больше g; но ведь невозможно найти такие три числа, для которых это было бы верно!

К счастью, в нашей задаче такие исключительные уравнения не возникают; во всяком случае, их не будет, если мы попытаемся провести кривую $y=gx^2+hx+k$ (параболу) через три точки (a, p), (b, q) и (c, r), где a, b и c различны[1]. Если же предположить, что $a=b$ или $a=c$, то такие неразрешимые уравнения могут появиться. Предположим, например, что мы пытаемся найти параболу для таблицы значений:

x	1	1	3
y	2	4	5

Это значит, что при $x=1$ выражение gx^2+hx+k должно равняться 2 и одновременно 4; но одно и то же выражение не может иметь два разных значения при $x=1$.

Это рассуждение показывает, что существуют системы уравнений, не имеющие решения; такие системы должны возникать всякий раз, когда наши требования нелепы, противоречивы. В этих случаях математика говорит свое «нет».

К счастью для научных работников, которым часто приходится строить кривые по точкам, полученным экспериментальным путем, неразрешимые уравнения встречаются только тогда, когда мы добиваемся чего-то явно несуразного. Но в нашей задаче нет бессмысленных ловушек: через любые три точки, соответствующие различным значениям x, действительно можно провести параболу. И так далее: через четыре точки можно провести кубическую параболу[2], через пять — параболу четвертого порядка, и так до бесконечности. Это можно доказать с помощью теории определителей, к которой мы еще не подошли.

Кроме доказательства того, что нужная функция существует, хорошо было бы знать, что она из себя представляет. Если нам удастся фактически найти функции, решающие рассмотренный тип задач, то автоматически отпадет необходимость доказывать, что такие функции существуют; это избавит нас от решения уравнений всякий раз, когда нам нужна такого рода функция, соответствующая экспериментальным данным.

[1] Это можно доказать с помощью определителей, о которых мы будем говорить в главе девятой.

[2] Т. е. кривую вида $y=ax^3+bx^2+cx+d$.— *Прим. ред.*

ПОИСКИ ЗАКОНОМЕРНОСТИ

Как же отыскивают функции, которые бы удовлетворяли нашим требованиям? В таких случаях самым разумным бывает *взять простейший пример и тщательно исследовать все, что он позволяет предположить для более сложных случаев.* Это общее правило; если вы не можете решить какую-нибудь задачу, составьте сами простейшую задачу того же типа и посмотрите, что она вам подскажет.

Для нашей задачи простейшим случаем является тот, с которого мы начали,— мы провели прямую $y=mx+c$ через две точки. Мы без всякого труда решаем приведенные выше уравнения (1) и (2) относительно m и c. Результат выглядит следующим образом:

$$m=\frac{p-q}{a-b},\quad c=\frac{aq-bp}{a-b},$$

а уравнение прямой имеет вид:

$$y=\left(\frac{p-q}{a-b}\right)x+\frac{aq-bp}{a-b}. \tag{3}$$

Указанные здесь алгебраические действия чрезвычайно просты. Но для читателя, не имеющего достаточно практики в алгебре, они могут представить некоторые затруднения. Такой читатель может пойти по одному из двух путей. Первый путь — принять полученный результат на веру и следить только за основным рассуждением, не вникая особенно в детали вычислений, с помощью которых получен результат. Впоследствии читатель, если он захочет, сможет проверить вычисления. Такой образ действия не утомляет мозг и позволяет ему легко справиться с основным доказательством.

Но есть люди, которые получают ощущение уверенности в том, что все понятно, лишь тогда, когда сами проделают все вычисления. Тогда им лучше всего и начать с вычислений. И если после этого читатель почувствует усталость, лучше отложить книгу на некоторое время и затем вернуться снова к основному доказательству со свежей головой.

Что мы можем узнать из уравнения (3)? Какие предположения оно позволяет нам сделать относительно общего результата?

На первый взгляд, не так уж много; приведенные выше алгебраические выражения выглядят довольно неприглядно.

Впрочем, мы замечаем, что $a-b$ — единственное выражение, стоящее внизу; оно является знаменателем дроби в выражениях для m и для c. И это не удивительно. Как мы ранее видели, при $a=b$ задача становится неразрешимой. Ее решение должно стать бессмысленным при $a=b$, и оно действительно становится таковым При $a=b$ все члены обращаются в бесконечность; решение приобретает бессмысленный вид $y=\infty x+\infty$.

Это наводит нас на мысль, что если мы имеем три значения x, а именно, a, b и c, то в знаменателе мы будем иметь только выражения $a—b$, $b—c$ и $c—a$.

Что касается остального, то довольно трудно предугадать, как будет выглядеть уравнение параболы, проведенной через три точки. Вероятно, если бы я занимался этим вопросом впервые, я попытался бы решить его прямыми алгебраическими вычислениями, а затем посмотрел бы, что можно почерпнуть из ответа. Вычисления были бы сложнее, а ответ, вероятно, еще более запутанным, но все-таки здесь можно было бы заметить некоторую закономерность.

На самом деле, одну важную черту этой функции мы уже нашли. Это удивительная черта, хотя я не стану упрекать вас, если вы ее не заметили.

Изучая алгебраическое выражение, следует всегда задавать себе вопрос: каким образом каждый отдельно взятый символ входит в него? Например, сложное выражение $b^2c^3+ac^5—7ab^4+bc$ является простым относительно a. Если считать все элементы, кроме a, постоянными, например, если положить $b=1$, $c=2$, то выражение упрощается до линейного выражения $25a+10$; и какие бы значения для b и c мы бы ни выбрали, выражение всегда будет линейно относительно a.

Нельзя сказать, чтобы уравнение (3) было особенно простым относительно a и b, но по отношению к p и q оно очень простое. Если мы зададимся значениями для a, b и x, скажем, $a=2$, $b=1$, $x=5$, то найдем, что $y=4p—3q$. И для любых других значений a, b, x мы всегда получим выражение вида $Ap+Bq$, линейное относительно p и q, где A и B — константы. Это нам подсказывает, что нужно преобразовать уравнение (3): сгруппировать члены, содержащие p, и члены, содержащие q. Проделав это, получим:

$$y=p\left(\frac{x-b}{a-b}\right)+q\left(\frac{a-x}{a-b}\right). \tag{4}$$

Наше выражение уже начинает оформляться. Чтобы понять его действительное содержание, лучше всего вернуться к исходной задаче. Мы искали функцию, для которой $y=p$ при $x=a$ и $y=q$ при $x=b$. Как же выражение (4) удовлетворяет этим требованиям?

Если мы положим $x=a$, выражение, стоящее в правой части, примет вид $p\cdot 1+q\cdot 0$. При $x=b$ получаем $p\cdot 0+q\cdot 1$.

Отсюда видно, как действует формула (4). Выражение (4) содержит и p и q. Но при $x=a$ y должно принимать значение p; q должно каким-то образом исчезнуть. Это достигается тем, что при $x=a$ множитель, стоящий в скобках при q, обращается в 0, и q не фигурирует в ответе. С другой стороны, при $x=b$ y должен обращаться в q. Поэтому при $x=b$ скобка перед q принимает значение 1.

Выражение, стоящее множителем при p, действует противоположным образом. При $x=b$ оно обращается в 0 и тем самым вычеркивает p из уравнения; при $x=a$ оно равно 1.

Обозначим соответствующие скобки через $f(x)$ и $\varphi(x)$. Тогда $y=pf(x)+q\varphi(x)$.

В новых обозначениях все, о чем мы только что говорили, выглядит следующим образом: $f(a)=1$, $f(b)=0$ — это значит, что p будет фигурировать в ответе при $x=a$ и отсутствовать при $x=b$. Равенства $\varphi(a)=0$, $\varphi(b)=1$ обеспечивают аналогичные условия для q, т. е. q не входит в ответ при $x=a$, но входит при $x=b$.

ПОСТРОЕНИЕ ТРЕБУЕМОЙ ФУНКЦИИ

Теперь в нашем распоряжении имеется идея, которая может быть применена к решению более общей задачи. Используем ее для нахождения параболы, удовлетворяющей условиям: $y=p$ при $x=a$, $y=q$ при $x=b$ и $y=r$ при $x=c$ [1].

Начнем с того, что положим $y=pu(x)+qv(x)+rw(x)$, где $u(x)$, $v(x)$, $w(x)$ — три функции, назначение которых состоит в том, чтобы p, q и r появлялись только тогда, когда это нужно. Например, p не должно входить в ответ ни при $x=b$, ни при $x=c$. Поэтому, когда x принимает значение b или c, функция $u(x)$ должна обращаться в нуль. Этого легко добиться. Если $u(x)$ содержит множитель $(x—b)$, то $u(x)=0$ при $x=b$. Если же $u(x)$ содержит $(x—c)$, то $u(0)=0$ при $x=c$. Поэтому мы предположим, что $u(x)$ содержит оба этих множителя. Далее, поскольку y должен быть многочленом только второй степени, то функции $u(x)$, $v(x)$ и $w(x)$ не должны содержать x в степени выше второй. Соответственно $u(x)$ может содержать только переменные множители $x—b$ и $x—c$. Но ничто не мешает, чтобы в $u(x)$ входил также и постоянный множитель. Поэтому можно предположить, что $u(x)=k(x—b)\ (x—c)$. Эта функция будет превосходно служить нашим целям: она будет исключать из выражения букву p при $x=b$ и при $x=c$ и будет оставлять одно только p при $x=a$.

Из условия $u(a)=1$ определяем значение k:

$$1=k(a-b)\ (a-c),$$

или

$$k=\frac{1}{(a-b)\ (a-c)}.$$

Как вы помните, мы так и предполагали: выражения $a-b$ и $a-c$ оказались в знаменателе. Итак,

$$u(x)=\frac{(x-b)\ (x-c)}{(a-b)\ (a-c)}.$$

[1] Здесь c, конечно, не имеет ничего общего с тем c, которое использовалось ранее в выражении $mx+c$.

Здесь выступает вполне отчетливая закономерность, поэтому без новых вычислений мы можем записать $v(x)$ в виде

$$\frac{(x-a)\,(x-c)}{(b-a)\,(b-c)}$$

и $w(x)$ в виде

$$\frac{(x-a)\,(x-b)}{(c-a)\,(c-b)}.$$

Собираем все воедино и получаем окончательный результат:

$$y=p\,\frac{(x-b)\,(x-c)}{(a-b)\,(a-c)}+q\,\frac{(x-a)\,(x-c)}{(b-a)\,(b-c)}+r\,\frac{(x-a)\,(x-b)}{(c-a)\,(c-b)}.$$

Придавая x поочередно значение a, b и c, вы можете убедиться, что эта формула действительно дает желаемые значения p, q и r.

Более того, если вы захотите перейти к следующей задаче — подбору формулы для кубической параболы, проходящей через четыре точки, вы убедитесь, что никаких новых идей не нужно. Закономерность, содержащаяся в приведенной выше формуле, легко обобщается. Формула для кубической параболы будет длиннее, но по существу получить ее будет не труднее, чем формулу для квадратной параболы.

Таким образом, о сложности алгебраического выражения нельзя судить по его длине или по количеству содержащихся в нем символов. Конечно, переписывать длинную формулу физически (да и умственно) утомительно; но если более длинная формула не приносит никаких новых идей, если ее закономерность уже распознана, ее не следует считать более сложной.

Мы могли бы, если бы понадобилось, написать формулу для параболы семнадцатой степени, проходящей через 18 заданных точек. Писать было бы довольно утомительно, но никаких новых идей не потребовалось бы.

ОБОБЩЕНИЕ И ИССЛЕДОВАНИЕ

Читая математические публикации, невольно поражаешься обилию работ, которые содержат попытку обобщить уже известный результат. Обобщение — это, вероятно, самый легкий и самый очевидный путь расширения математических знаний.

Разумеется, самое естественное — это взять какую-нибудь полезную задачу и попытаться решить ее; многие исследования начинаются с прямой попытки решить задачу. Но по-настоящему трудные задачи редко поддаются атаке в лоб. Можно часами ломать себе голову и так и не придумать, с чего начать; зачастую даже и представить себе не можешь, как выглядит решение. Воображение нуждается в пище, идеи не возникают из ничего.

Поэтому часто пытаешься начать с конца. Анализируешь метод, оказавшийся удачным в прошлом, пытаешься обобщить его и, наконец, видишь, какие именно задачи можно решить с помощью обобщенного метода.

Часто, конечно, желание решить задачу воздействует на выбор темы; пытаешься обобщить методы, которые с успехом применялись для решения более простых задач того же типа.

Элемент обобщения содержится в каждом новом открытии. Человек не может не находиться под влиянием имеющихся у него знаний; новое открытие вырастает из старых знаний.

Иногда, конечно, пытаясь обобщить известное, человек сталкивается с совершенно новым и неожиданным. Он может даже обнаружить, что обобщение невозможно, что старый способ или результат является своеобразным, уникальным.

Это замечание может быть проиллюстрировано на примере, рассмотренном в главе третьей. Мы нашли, что $(ax+by)^2$ никогда не превосходит $(a^2+b^2)\ (x^2+y^2)$ для действительных значений символов a, b, x, y. Этот одиночный результат может привести к исследованиям в нескольких направлениях.

В одном направлении этот результат можно обобщить до бесконечности. Так, мы найдем, что $(ax+by+cz)^2$ никогда не превосходит $(a^2+b^2+c^2)\ (x^2+y^2+z^2)$ и подобный же результат можно получить для четырех, пяти, шести и вообще любого желаемого количества членов, стоящих в скобках.

Можно также попытаться, вместо того чтобы обобщать результат, обобщить сам метод доказательства. Предложение, что одна функция никогда не превосходит другую, называется, как известно, неравенством. Неравенство, приведенное в главе третьей, мы доказали, установив, что разность двух функций представляет собой полный квадрат. Можно ли все неравенства доказывать таким способом? Ответ утвердительный. Если одно выражение меньше другого для любых действительных значений символов, то разность между ними можно представить в виде суммы квадратов.

В главе третьей мы рассматривали тождество

$$(a^2+b^2)\ (x^2+y^2)=(ax+by)^2+(bx-ay)^2.$$

Это тождество выявляет замечательную закономерность, которая уводит нас совсем в другом направлении. Первая скобка a^2+b^2 есть сумма двух квадратов. Вторая скобка x^2+y^2 — тоже сумма двух квадратов. Приведенное выше тождество формулируется следующим образом: «Произведение суммы двух квадратов на сумму двух квадратов является суммой двух квадратов».

Можно ли это обобщить? Можно ли в этой формулировке заменить слово «два» любым другим числом?

Конечно, мы можем заменить его словом «один», так как

$$(a)^2\ (x)^2=(ax)^2.$$

Однако в этом результате радости мало. Гораздо внушительнее формула выглядит при замене слова «два» словом «четыре». Вот ее вид:

$$(a^2+b^2+c^2+d^2)(x^2+y^2+z^2+t^2)=$$
$$=(ax+by+cz+dt)^2+(bx-ay+dz-ct)^2+$$
$$+(cx-dy-az+bt)^2+(dx+cy-bz-at)^2.$$

Можно было бы написать и тождество, где произведение восьми квадратов на восемь квадратов равно сумме снова восьми квадратов. Я не стану воспроизводить его здесь.

Итак, мы получили тождества для 1, 2, 4 и 8 квадратов. Каким же будет следующее? Думается, что это будет тождество для 16 квадратов. Однако это не так; последовательность здесь прерывается. Было доказано, что такого рода тождества существуют только для 1, 2, 4 и 8. Здесь мы имеем нечто не поддающееся обобщению.

АЛГЕБРА

В тесной связи с только что рассмотренным вопросом находится предмет линейной алгебры. Обычно люди вполне удовлетворяются действительными числами, вроде 5 или $\frac{3}{4}$, но электротехники и математики знают, что большое удобство дает введение нового символа i, такого, что $i^2=1$. Тогда со сложным выражением $x+iy$, называемым комплексным числом, можно обращаться как с обычными алгебраическими величинами.

Естественно спросить: «Если введение нового символа оказалось полезным, отчего же не ввести еще несколько новых букв и не посмотреть, не приведет ли это к хорошим результатам?»

Это направление исследований звучит довольно обещающе, но ответ оказывается отрицательным. Доказано, что расширение области действительных чисел в сторону комплексных чисел вида $x+iy$— это единственное расширение, которое можно сделать, сохраняя правила элементарной алгебры. Комплексные числа, таким образом, не просто полезны; они являются единственными в своем роде, уникальными.

Чтобы пойти дальше комплексных чисел, нужно приготовиться кое-чем пожертвовать. Обычно приходится поступиться законом $ab=ba$. Если мы не возражаем, чтобы a, умноженное на b, не было обязательно равно $b\cdot a$, мы можем перейти к кватернионам, где мы встретим такие числа, как, скажем, $3+4i+5j+6k$. Правила перемножения букв i, j, k гораздо более сложны, чем правила для обычных комплексных чисел. Они выражаются следующими равенствами:

$$i^2=-1,\quad j^2=-1,\quad k^2=-1;$$
$$ij=k,\qquad ji=-k;$$
$$jk=i,\qquad kj=-i;$$
$$ki=j,\qquad ik=-j.$$

Я привожу эти равенства на тот случай, если вам захочется использовать их для получения тождества для четырех квадратов, приведенного в предыдущем разделе.

Если вы умножите $(a+ib+jc+kd)$ на $(x-iy-jz-kt)$ по правилам для кватернионов, то окажется, что в произведении члены, содержащие i^2, ij, ..., jk, k^2, взаимно уничтожатся и все произведение будет иметь вид:

$$(\quad)+i(\quad)+j(\quad)+k(\quad),$$

где в скобках содержатся некоторые алгебраические выражения. Эти выражения идентичны с соответствующими выражениями в правой части тождества для четырех квадратов.

Кватернионы сами по себе являются вехой на пути к дальнейшему расширению наших алгебраических познаний. Если вы хотите пойти еще дальше, вам придется приготовиться к дальнейшим жертвам и поступиться рядом обычных законов алгебры. Вы должны быть готовы к тому, что встретите числа p и q, отличные от нуля, которые тем не менее в произведении дают нуль! Такое явление совершенно невозможно в обычной арифметике.

Действительные числа, комплексные числа и кватернионы отделены от остальной алгебры огромной пропастью и представляют собой отдельный класс, являющийся предметом математического изучения. Все они имеют как практическое, так и теоретическое значение.

Глава пятая

ОБ УНИФИКАЦИИ

В течение всей своей жизни он искал скрытую взаимосвязь вещей и лежащее в их основе единое начало. (О Фридрихе Фрёбеле)

Р. Х. Квик,
Реформаторы образования

В предыдущей главе мы увидели, что каждый отдельный результат может быть осмыслен как источник обобщений, которые могут иметь разные направления. Мозг принужден рассматривать эти обобщения и в то же самое время чувствовать тяжелое бремя этого нескончаемого разнообразия. Становится совершенно необходимым иметь знания до приемлемых пределов, объединять разрозненные результаты.

Один из самых приятных моментов в истории математики — это момент, когда выясняется, что два раздела математики, которые ранее рассматривались отдельно и считались несвязанными, в действительности являются двумя скрытыми формами одного и того же.

Один из таких поразительных примеров унификации мы находим в пределах школьной программы. Школьная математика распадается на две части: с одной стороны, арифметика, из которой затем развивается алгебра; и та и другая оперируют числами, величинами; с другой стороны, геометрия и ее продолжение — тригонометрия, по необходимости использующая числа, но, главным образом, занимающаяся геометрическими формами. Тригонометрия использует алгебраические приемы, но основа ее, во всяком случае, лежит в геометрии; эта основа весьма отлична от счета, являющегося основой арифметики.

Но вот некоторые определенные закономерности начинают проявляться в обоих этих разделах. В алгебре, возводя $(1+x)$ в различные степени, мы находим:

$$1 + x = 1 + x,$$
$$(1 + x)^2 = 1 + 2x + x^2,$$
$$(1 + x)^3 = 1 + 3x + 3x^2 + x^3,$$
$$(1 + x)^4 = 1 + 4x + 6x^2 + 4x^3 + x^4, \ \ldots$$

и таким образом получаем числа треугольника Паскаля:

```
      1   1
    1   2   1
  1   3   3   1
1   4   6   4   1
```

С другой стороны, в тригонометрии мы встречаем формулу для тангенса суммы двух углов, а именно:

$$\operatorname{tg}(A+B)=\frac{\operatorname{tg}A+\operatorname{tg}B}{1-\operatorname{tg}A\cdot\operatorname{tg}B}.$$

Из этой формулы можно найти $\operatorname{tg}2A$, полагая $A=B$. Затем, положив $B=2A$ и используя предыдущий результат, можно найти $\operatorname{tg}3A$ и т. д. Если для краткости мы обозначим $\operatorname{tg}A=t$, то результат будет иметь вид:

$$\operatorname{tg}A=t,$$

$$\operatorname{tg}2A=\frac{2t}{1-t^2},$$

$$\operatorname{tg}3A=\frac{3t-t^3}{1-3t^2},$$

$$\operatorname{tg}4A=\frac{4t-4t^3}{1-6t^2+t^4}\ \ldots$$

Здесь появляется тот же самый набор чисел. Например, числа 1, 4, 6, 4, 1, встречавшиеся в выражении $(1+x)^4$, появляются в формуле для $\operatorname{tg}4A$, но по зигзагообразной линии:

$$\begin{matrix} & 4 & & 4 & \\ 1 & & 6 & & 1 \end{matrix}$$

Конечно, перед некоторыми из них стоит знак минус, но тем не менее тот факт, что здесь встречаются числа треугольника Паскаля [1], является достаточным свидетельством общей закономерности, лежащей в основе обеих наук.

Мы могли бы отметить это просто как забавное совпадение. Но в главе второй мы подчеркивали, что закономерность очень важна, она является признаком какой-то важной взаимосвязи и требует внимательного изучения.

Если бы мы пожелали узнать основную причину этой общей закономерности, мы пришли бы к знаменитому равенству:

$$e^{i\theta}=\cos\theta+i\sin\theta.$$

Это соотношение означает полное присоединение тригонометрии к алгебре. Мы можем теперь определить $\cos\theta$ и $\sin\theta$ следующим образом:

$$\cos\theta=\frac{1}{2}(e^{i\theta}+e^{-i\theta})\ \text{ и }\ \sin\theta=\frac{1}{2}(e^{i\theta}-e^{-i\theta}),$$

т. е. дать чисто алгебраические определения, совсем не обращаясь к геометрии.

[1] По поводу треугольника Паскаля см.: Е. Б. Дынкин и В. А. Успенский. Математические беседы. Гостехиздат, 1952.— *Прим. ред.*

Тригонометрия, таким образом, превращается просто в раздел алгебры. Все свойства синусов и косинусов следуют из приведенного выше определения и могут быть доказаны гораздо быстрее и проще, чем обычными элементарными методами.

Искусный учитель может подвести учеников к тому, что они сами «откроют» бином Ньютона и отношение тригонометрии к алгебре, воспроизведя, таким образом, в классе великие открытия XVII и XVIII веков и доставив классу удовольствие присутствовать при математическом открытии и участвовать в нем.

Самое большое удовольствие от такого унифицирующего открытия получит тот, кто долго боролся с огромной массой неусвоенной информации в старой форме и кто хорошо знаком со всем объемом материала. Например, ученики, проработавшие школьный курс тригонометрии, оценят по достоинству объединение тригонометрии и алгебры, о котором мы рассказываем в этом разделе. Но такого удовлетворения не получит человек, которому преподносят много новых сведений, понятных ему лишь отчасти, и который вдруг узнает о возможности унификации. По этой причине многие примеры математической унификации, которые были бы интересны студенту-математику, не подходят для обсуждения в этой главе.

ГИПЕРГЕОМЕТРИЧЕСКИЕ ФУНКЦИИ

В конце второй главы упоминался замечательный пример унификации: функция, которая содержит в себе почти все ранее изученные функции.

Имеется несколько различных точек зрения, с которых можно рассматривать гипергеометрические функции. Здесь мы остановимся только на одной из них. Она не является самой поучительной, но зато описание ее не займет много времени. Это — представление гипергеометрической функции с помощью ряда.

Очень многие функции обладают тем свойством, что их можно разложить в ряд. В учебниках вы найдете такие примеры разложения в ряд, как

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

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

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

$$\operatorname{arc\,tg} x = x - \frac{1}{3}x^3 + \frac{1}{5}x^5 - \frac{1}{7}x^7 + \ldots,$$

$$e^x = 1 + \frac{x}{1} + \frac{x^2}{1\cdot 2} + \frac{x^3}{1\cdot 2\cdot 3} + \frac{x^4}{1\cdot 2\cdot 3\cdot 4} + \ldots,$$

не говоря уже о более внушительных результатах, вроде

$$\frac{1}{2}(\arcsin x)^2 = \frac{x^2}{2} + \frac{2}{3}\cdot\frac{x^4}{4} + \frac{2\cdot4}{3\cdot5}\cdot\frac{x^6}{6} + \frac{2\cdot4\cdot6}{3\cdot5\cdot7}\cdot\frac{x^8}{8} + \ldots$$

Это лишь несколько примеров рядов, обладающих следующими двумя свойствами: (I) функция, которую представляет этот ряд, встречается естественным образом в элементарной математике; (II) ряд характеризуется определенной закономерностью, которая позволяет написать любой из следующих членов ряда. Например, пятый член разложения в ряд функции arc tg x будет, очевидно, $+\frac{1}{9}x^9$, а следующий за написанными член в последнем, довольно сложном, на первый взгляд, разложении, будет

$$\frac{2\cdot4\cdot6\cdot8}{3\cdot5\cdot7\cdot9}\cdot\frac{x^{10}}{10}.$$

Всякий, кому захочется ознакомиться с бо́льшим количеством таких примеров, найдет их в задачниках по дифференциальному исчислению.

При изучении более сложных примеров таких рядов напрашивается вывод, что все они прямо или косвенно связаны со следующим рядом:

$$1 + \frac{a\cdot b}{1\cdot c}x + \frac{a(a+1)\cdot b(b+1)}{1\cdot 2c(c+1)}x^2 + \frac{a(a+1)(a+2)\cdot b(b+1)(b+2)}{1\cdot2\cdot3c(c+1)(c+2)}x^3 + \ldots$$

Обозначим этот ряд для краткости через $F(a, b, c; x)$, так как он содержит три постоянные a, b, c, которым мы можем придать любые значения, и переменную x. Закон, по которому построен этот ряд, вовсе не сложен; например, вы без всякого труда напишете член ряда, содержащий x^4 или x^5.

Некоторые из ранее рассмотренных рядов могут быть получены непосредственно из последнего ряда. Возьмем самый простой случай: если мы положим $a=b=c=1$, то получим $1+x+x^2+x^3+\ldots$, т. е. первый пример из нашего списка. Второй ряд мы получим, положив $a=2, b=1, c=1$.

Очевидно, что нельзя получить третий ряд, задавшись некоторыми значениями величин a, b и c, так как он начинается с x, в то время как $F(a, b, c; x)$ начинается с 1. Однако, приняв $a=1$, $b=1$, $c=2$, мы получаем ряд, близкий к тому ряду, который нас интересует. В каждом члене не хватает лишь множителя x, и нет необходимой последовательности знаков + и —. Но эти дефекты быстро устраняются — легко проверить, что

$$\ln(1+x) = xF(1,\ 1,\ 2;\ -x).$$

Для функции arc tg x, где встречаются только нечетные степени x, формула принимает вид:

$$\operatorname{arc\,tg} x = xF\left(\frac{1}{2},\ 1,\ 1\frac{1}{2};\ -x^2\right).$$

Чтобы получить ряд для e^x, мы должны допустить, что некоторые из величин стремятся к ∞.

Таким образом, применяя несложные операции: введение дополнительного множителя x, замену x на $-x$ или на $-x^2$, а также придавая константам различные значения, мы получаем из $F(a, b, c; x)$ разложение в ряд почти любой элементарной функции. Функция $F(a, b, c; x)$ оказывается замечательно приспособляемой.

Кроме функций, встречающихся в школьном курсе, имеется еще ряд функций, используемых инженерами или физиками — сюда относятся, например, полиномы Лежандра или бесселевы функции, которые также являются частными случаями гипергеометрической функции. Пожалуй, во многих университетах в настоящее время 95% — если не все 100% — функций, изучаемых студентами-физиками, инженерами и даже математиками, могут быть представлены одной-единственной формулой

$$F(a, b, c; x).$$

Значит ли это, что не существует других функций, кроме функций гипергеометрического типа? Конечно, нет; очень легко привести примеры функций другого типа. Объяснение совсем в другом.

ОГРАНИЧЕННОСТЬ НАШИХ ПОЗНАНИЙ

Вообразите себе фермеров, живущих в стране, где не знают другого орудия труда, кроме деревянного плуга. По необходимости фермы должны располагаться в тех местах, где земля достаточно мягка, чтобы ее обрабатывать деревянным инструментом. Если население увеличится настолько, что сможет заселить всю имеющуюся мягкую землю, то расположение ферм будет представлять собой карту районов с мягкой почвой. Каждый, кто посмеет переступить эту область, обречен на гибель.

В этом примере имеется много общего с математическими исследованиями. На какой-то стадии истории математики обладают некоторыми запасами знаний, опыта, воображения. Этих запасов достаточно, чтобы решить одни задачи, но они не дают возможности решать другие. Если математик занимается задачей, которая лежит за пределами имеющихся в его распоряжении сведений, он не публикует никаких результатов и не оставляет следа в истории математики. Другие же математики, занимающиеся доступными им задачами, публикуют открытия. Этот процесс происходит бессознательно, поэтому карта математических знаний становится похожей на карту задач, разрешимых с помощью имеющихся средств.

Но, конечно, сами публикуемые открытия приводят к изобретению новых средств. Так же как приход стального плуга изменил карту сельскохозяйственных поселений, так и эти новые средства открывают новые области для плодотворных исследований. Однако могут понадобиться века, чтобы эти новые средства появились;

и пока мы их ждем, граница наших познаний остается непроходимой преградой.

В случае с гипергеометрической функцией мы, кажется, имеем нечто в этом роде. Она, по-видимому, является пределом для того типа закономерностей, которые мы сейчас знаем. Если кто-либо попытается проникнуть за эти пределы, все ему кажется бесформенным. Несомненно, закономерности существуют и там, но это те закономерности, которые мы еще не научились видеть [1].

Я не хотел бы утверждать, что гипергеометрическая функция — это единственная функция, которая в какой-то мере известна математикам. Это далеко не так. Имеются и другие плодородные долины, которые можно обрабатывать деревянным плугом двадцатого века. Но долина, за пределы которой еще не заглядывали школьники, инженеры, физики и все изучающие элементарную математику,— это долина гипергеометрической функции, а ее границы (за исключением одной или двух трещин, изученных исследователями) — нетронутые скалы.

---

[1] В последнее десятилетие развитие одной из алгебраических теорий (теории представлений групп) привело к значительному обобщению понятия гипергеометрической функции.— *Прим. ред.*

Глава шестая

НЕЕВКЛИДОВЫ ГЕОМЕТРИИ

Пусть обиталищем моим была бы скорлупа ореха,
Я все ж владыкой необъятного себя бы мнил.

Гамлет[1]

НАУЧИТЬСЯ ЗАБЫВАТЬ

Развитие современной математики затрудняется не тем, что трудно освоиться с новыми идеями, а тем, что трудно отказаться от старых. Возьмите, например, теорию относительности Эйнштейна. Я не думаю, чтобы ангелу — существу, лишенному телесной оболочки, не имеющему ни малейшего представления о пространстве и времени,— было бы труднее понять теорию Эйнштейна, чем старые взгляды на Вселенную. Новую теорию часто трудно понять потому, что человеку свойственно сохранять образ мысли, связанный со старой теорией. В течение последних 150 лет математические представления претерпевают постоянное изменение. Традиционная математика подвергалась тщательной ревизии, и оказалось, что отдельные идеи бессмысленны или неверны. Если математическое образование и представляет какую-то ценность как средство, оказывающее влияние на формирование общего мировоззрения, то эта ценность, по всей вероятности, заключается в постоянной тренировке ума. Нынешний век — ярчайший пример хаоса, возникшего в результате медленного движения человеческих идей. Очень вероятно, что будущим поколениям,— если, конечно, такие будут, если мы не уничтожим в своих блужданиях в потемках основы жизни,— мы покажемся дикарями и сумасшедшими. Но нам трудно правильно оценить современный мир; нам мешает туман прошлого.

Неевклидова геометрия — одна из новых идей в математике; это первая и самая крупная ломка традиций. До 1800 года,— а я думаю, что для многих школьников и школьниц и по сей день,— евклидова геометрия оставалась и остается единственно правильной, она казалась чем-то определенным и установленным раз и навсегда. Как дети, мы склонны думать о какой-то определенной «Геометрии», а не вообще о «геометрии».

ФИЗИЧЕСКИЕ И МАТЕМАТИЧЕСКИЕ ИСТИНЫ

Необходимо строго разграничить то, что истинно в физическом смысле, от того, что истинно математически. Физика основывается на эксперименте. Физической истиной, например, является то, что

[1] Перевод Б. Л. Пастернака.— *Прим. перев.*

Земля представляет собой (приближенно) шар радиусом 6360 *км*, находящийся на расстоянии 150 000 000 *км* от Солнца. Но очень легко представить себе и другое. Нетрудно вообразить, например, что Земля в действительности больше или меньше или имеет форму куба или тетраэдра; нет особых резонов и для того, чтобы ее расстояние от Солнца равнялось именно 150 миллионам километров, а не 80 или 200 миллионам.

Математика имеет дело только с логической связью между понятиями. Ее интересует, как одно следует из другого. Возьмем пример из другой области. Суд присяжных иногда может вынести приговор, который, являясь логически правильным, по существу неверен. Предположим, что подсудимый совершил убийство, но имеющихся свидетельских показаний недостаточно, чтобы доказать его вину. Присяжные поступают правильно, вынося приговор «Не виновен». Вина не была *доказана,* другой вопрос — существует ли она. В этом случае присяжные поступают подобно математикам. И если после суда появятся новые улики, то судей винить нельзя. Точно так же можно сказать, что генерал действовал разумно, исходя из сведений, имевшихся у него во время сражения, даже если действительные последствия его действий оказались гибельными ввиду ряда обстоятельств, о которых он не знал и не мог знать.

Поэтому математики не ставят вопроса, равна ли в действительности сумма углов треугольника 180°, их интересует, есть ли *логическая необходимость* в том, чтобы сумма углов треугольника равнялась 180°. Если бы евклидова геометрия была логически единственно возможной, то это означало бы, что если вы намерены создать вселенную, она должна быть построена по законам евклидовой геометрии. Можем ли мы представить себе вселенную с геометрией, отличной от нашей?

Для физика вопрос будет состоять в другом: «Является ли геометрия нашей Вселенной действительно евклидовой?»

С точки зрения математики определенно возможны и другие геометрии; с точки зрения физики может оказаться, что геометрия Евклида не совсем верна для нашей Вселенной.

ЧИСЛО ИЗМЕРЕНИЙ

Мы можем начать расшатывать наши представления с вопроса о числе измерений. Мы живем в трехмерном пространстве: вы можете пройти x *км* на восток, y *км* на север и подняться на z *км* вверх. Вы окажетесь на расстоянии s *км* от исходной точки, где s определяется из равенства

$$s^2=x^2+y^2+z^2.$$

А если у вас нет воздушного шара или вертолета и вы не можете оторваться от поверхности земли, то вы оказываетесь заключенным в двумерном пространстве. Вы можете сделать x *км* на восток и y *км* на север, и пройденное вами расстояние s определится из равенства

$$s^2 = x^2 + y^2.$$

А если мы еще больше ограничим вас, потребовав, чтобы вы ехали в поезде, пересекающем страну с запада на восток, тогда вы будете двигаться только в одном направлении. Вы проедете x *км* на восток (если ваш поезд мчится с востока на запад, то x окажется отрицательным числом), и проделанный вами путь s определится из равенства $s^2 = x^2$, которое я оставляю в такой форме, чтобы сохранить аналогию с двумя предыдущими формулами. Наконец, если вас привязать к столбу, то вы не смогли бы двигаться вообще; вы оказались бы в пространстве 0 измерений.

Нам кажется совершенно естественным существование четырех возможных пространств: точек, линий, плоскостей и объемных тел. Но давайте-ка посмотрим, как все это выглядит с точки зрения жителя какой-нибудь другой вселенной. Представьте себе, что вам нужно по телефону объяснить идеи геометрии ангелу. Когда я говорю «ангел», я имею в виду существо, лишенное какого бы то ни было физического опыта. «Длина», «цвет» и прочие понятия ему ни о чем не говорят. Но чтобы мы могли общаться с ним, нам придется наделить его способностью слышать и говорить и предположить, что он чрезвычайно понятлив.

Мы, конечно, лишены возможности показывать ему чертежи или наброски, во-первых, потому, что мы разговариваем с ним по телефону, а во-вторых, потому, что он не способен видеть или осязать. Но мы можем ему передать значение цифр. Мы стучим три раза по микрофону и говорим: «это были три удара». Так как ангел очень понятлив, он, конечно, быстро получает представление о целых числах 1, 2, 3, 4, 5 и т. д., а, пользуясь ими, мы объясняем ему дроби, скажем, $\frac{3}{5}$ или $\frac{2}{7}$ и так далее; нам также удается дать ему представление об иррациональных числах, вроде $\sqrt{2}$, π и e. Все эти понятия объясняются, конечно, чисто арифметически. Мы не можем, например, объяснить ему, что π — это отношение длины окружности к диаметру, потому что ангел не знает, что такое окружность; но зато мы можем представить ему число π в виде бесконечного ряда.

Перейдем теперь к геометрии. Очевидно, у нас есть только один способ объяснить ему геометрию — при помощи аналитической геометрии, где любые геометрические положения получают свое арифметическое и алгебраическое выражение. Итак, начнем. Ангел не знает, что означают слова «восток», «север», «вверх», поэтому мы не можем воспользоваться этими понятиями. Мы начинаем: «Прямая состоит из точек. Вы не можете понять, что мы подразумеваем под словом «прямая» или «точка», но так или иначе положение точки на прямой определяется с помощью единственного числа x.

На плоскости точка определяется парой чисел (x, y). В трехмерном пространстве для определения положения точки нужны три числа (x, y, z). Затем имеется еще расстояние. Оно измеряется числом s:

на прямой	$s^2 = x^2$;
на плоскости	$s^2 = x^2 + y^2$;
в пространстве	$s^2 = x^2 + y^2 + z^2$».

Здесь мы останавливаемся. Ангел разочарован. Он ожидал, что мы пойдем дальше и скажем, что в четырехмерном пространстве точка определяется четырьмя числами x, y, z, t, а расстояние s — по формуле $s^2=x^2+y^2+z^2+t^2$; затем он надеялся услышать о пяти, шести, семи измерениях и так до бесконечности.

С точки зрения математики нет никаких оснований для того, чтобы остановиться на числе 3, а не на каком-нибудь другом числе. В нашей вселенной «север» перпендикулярен «востоку» и «вверх» перпендикулярен обоим этим направлениям; но мы не можем найти четвертое направление, которое было бы перпендикулярно всем этим трем. И все же нет причин для того, чтобы не существовало вселенной четырех, пяти или шести измерений. Мы привыкли к трем измерениям, но это тем не менее не причина.

Конечно, нет необходимости доходить и до 3. Можно было бы создать вполне удобную вселенную только в двух измерениях. Людям пришлось бы перемещаться только в плоскости. В такой вселенной существовали бы только направления на север и восток, но не было бы направления «вверх». Вы можете спросить: «А что же находится над или под плоскостью?» Ответа нет; это вопрос, который нельзя задавать. Он не имеет физического смысла в плоской вселенной. Житель трехмерного пространства с тем же успехом мог бы спросить нас: «Вы можете перемещаться на север, на восток и вверх, а что случится, если вы захотите пойти в четвертом направлении, под прямым углом ко всем трем?». Единственное, что мы можем ответить, это следующее: «Не существует направления, перпендикулярного всем трем».

Позже в этой главе мы будем говорить о существах, живущих на поверхности шара. Вы захотите спросить: «Что находится внутри шара и что снаружи?». Но вы не должны так спрашивать. Поверхность шара является их вселенной. Для них все остальное физически нереально.

В самом деле, сама природа вселенной изменилась бы, если существа, ее населяющие, получили бы возможность войти в лишнее измерение, т. е. если идея еще одного измерения стала бы физической реальностью.

Пусть, например, на плоскости (рис. 9) квадрат представляет собой тюрьму. Если нарисовать квадрат на крышке стола и положить внутрь него монету, то в двумерной плоскости монета может покинуть квадрат, только пересекая его стенки. Но если допустить

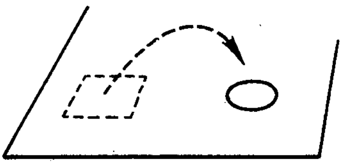


Рис. 9.

движение в третьем направлении — вверх, то монета сможет выйти из квадрата, не пересекая его стен.

Точно так же, если в трехмерном пространстве мы получили бы возможность двигаться в четвертом направлении, мы могли бы выйти из закрытой тюремной камеры. Вообразите, например, что в вашей власти было бы совершить путешествие во времени. Если вы были бы заперты в крепости, вы могли бы вернуться назад к тому времени, когда крепость еще не была построена, пройтись полмили и затем вернуться в сегодня. Для нас же, обычных обитателей трехмерного пространства, это выглядело бы, как если бы вы исчезли из крепости и очутились в полумиле в стороне от нее. Но это чистая фантастика. Вообразить мы это можем, но сделать — нет. Математически это реально, физически — немыслимо. Точно так же, если мы рассматриваем воображаемую вселенную, которая является поверхностью шара, мы должны помнить, что для ее обитателей только эта поверхность физически реальна. Для нас, как для создателей этой вселенной, возможно, имеет смысл говорить о том, что находится внутри шара или вне его; философы нашей сферической вселенной могут размышлять о таких вещах подобно тому, как я рассуждал о путешествии во времени, но это лишь бесполезные рассуждения, ведь ни одно из наших созданий, обитателей двумерного мира, не могло бы покинуть поверхность шара. Это их мир, это их реальность.

Следует внести ясность, в каком смысле математики теперь используют термин «неевклидова геометрия». Его употребление вовсе не связано с числом измерений изучаемой вселенной. Правда, вселенная 4, 5 или 6 измерений сильно отличалась бы от той Вселенной, которую имел в виду Евклид, но термин «неевклидова» имеет другое содержание. Основная теорема евклидовой геометрии — теорема Пифагора. На плоскости она выражается формулой $s^2=x^2+y^2$, которую мы уже приводили раньше. Для трехмерного пространства формула принимает вид $s^2=x^2+y^2+z^2$,
для пространства четырех измерений $s^2=x^2+y^2+z^2+t^2$,
для пяти $s^2=x^2+y^2+z^2+t^2+u^2$ и т. д.

Геометрии этих пространств нужно соответственно называть: «евклидово четырехмерное пространство» и «евклидово пятимерное пространство», поскольку в их основе лежит теорема Пифагора в несколько измененной форме. А под термином «неевклидова геометрия» мы имеем в виду такую геометрию, для которой теорема Пифагора больше не действительна. Это более глубокое изменение, чем простое добавление одного-двух измерений.

МАТЕРИАЛЫ ДЛЯ НОВОЙ ВСЕЛЕННОЙ

Возможно, математики зря тратят время, обдумывая, какой *могла бы* быть Вселенная, вместо того, чтобы быть хорошими физиками и обсуждать, какая она *есть*. Луна могла бы быть сделана из зеленого сыра, но это не имеет ни малейшего значения. Точно так же можно согласиться, что совершенно бесполезно обсуждать, может ли Вселенная иметь четыре или только два измерения, когда на самом деле у нее их три.

В одном рассказе Грэма Грина адвокат задает вопрос свидетелю: «Вы уверены, что это как раз тот человек, которого вы видели? — «Да».— «Могли бы вы поклясться, что это не был его брат-близнец, сидящий в зале суда?». После того как свидетель взглянул на близнеца, он признал, что их совершенно невозможно различить.

Математики сделали нечто подобное со вселенной. Мы были готовы поклясться, что живем во Вселенной № 1 — евклидовой Вселенной. Математики создали еще две вселенные: Вселенную № 2 и Вселенную № 3. Мы находимся в положении ребенка, который выбежал на улицу, застроенную типовыми домами, и никак не может вспомнить, в каком доме он живет.

Давайте сначала отметим, в чем сходство между Вселенными 1, 2 и 3. Во всех Вселенных существуют твердые тела, способные двигаться. Когда я говорю «твердое тело», я имею в виду нечто вроде кирпича или стального бруска, а не глину или замазку. Вы не станете делать линейку из замазки или эластичного материала.

Очень просто проверить, является ли тело «твердым». На рис. 10 точки A и B подковообразного предмета касаются точек C и D, отмеченных на линейке. Я разделяю предметы, затем снова их сближаю. И если я нахожу, что сколько бы раз я ни проделал это, я могу совместить точки A и B с точками C и D, я заключаю, что оба тела — твердые. Тогда я могу ввести понятие *длины* и сказать, что длина отрезка AB равна длине отрезка CD.

Далее, мы предполагаем, что в каждой из Вселенных 1, 2 и 3 тела ведут себя одинаково в любом месте. Если $AB = CD$, когда я сравниваю предметы здесь, то, отвези я их в Америку и измерь там, они были бы тоже одинаковой длины.

До сих пор мы сделали два предположения о сходстве всех Вселенных: 1) *твердые тела, свободно подвижные, позволяют определять длину; 2) пространство обладает одинаковыми свойствами во всех своих точках.*

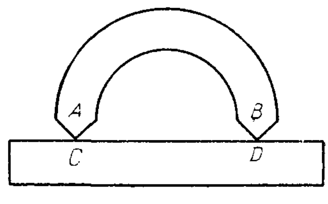


Рис. 10.

Прямой мы будем называть кратчайшее расстояние между двумя точками. Физически мы могли бы составить прямую из цепи жестких тел, как показано на рис. 11. Если мы до предела растянем точки A и E, то точки

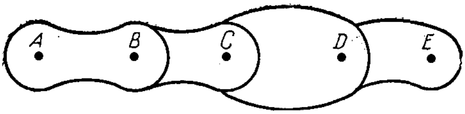


Рис. 11.

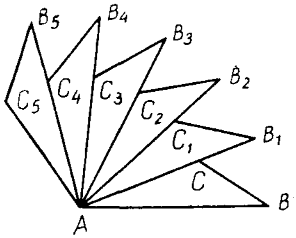


Рис. 12.

B, *C* и *D*, являющиеся местами соединения отдельных звеньев цепи, будут лежать на одной прямой.

Сделаем еще одно предположение: 3) *существует только одна прямая, соединяющая две данные точки.*

Вы можете заметить, что твердые тела позволяют нам измерять не только длину, но и угол. На рис. 12 показано, как можно сделать транспортир. Берем жесткий треугольник ABC. Вращением вокруг точки A приводим его в положение AB_1C_1, где направление AB_1 совпадает с первоначальным направлением AC. Затем поворачиваем наш треугольник до положения AB_2C_2 и так далее. Если у треугольника ABC угол A достаточно острый, то отрезки AB, AB_1, AB_2 и т. д. будут лежать очень близко друг от друга, т. е. мы сможем измерять углы с большой точностью. В привычной для нас Вселенной 1 существуют прямые углы. Точно так же можно говорить о прямых углах применительно ко Вселенной 2 и Вселенной 3. Можно нарисовать и круги. Если все точки B_1, B_2, B_3... лежат на одинаковом расстоянии от точки A, то это означает, что они лежат на окружности с центром в точке A. Все это нам знакомо и привычно.

ЧТО В ЕВКЛИДОВОЙ ГЕОМЕТРИИ ПОКАЗАЛОСЬ СОМНИТЕЛЬНЫМ?

Большинство из того, что Евклид предполагает или считает само собой разумеющимся, действительно очень «разумно», если это рассматривать, как физическое описание нашей Вселенной. Обычный человек готов согласиться с этим. Но есть у Евклида одно предположение, которое не так уж очевидно. Он предполагает следующее: через точку A (рис. 13) можно провести одну и только одну прямую AB, параллельную CD, причем сумма углов BAC и ACD равна 180°. Если их сумма меньше 180°, то AB пересечет CD где-то правее точки C. Если же их сумма больше 180°, то продолжение прямой AB пересечет прямую CD левее точки C. Большинство из нас принимают это как теорему, как нечто

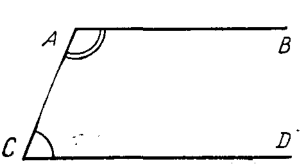


Рис. 13.

доказанное с помощью других более простых предположений. Но Евклид не дает это утверждение как следствие других идей; вы должны принять его без доказательств с самого начала, без этого нельзя пойти дальше.

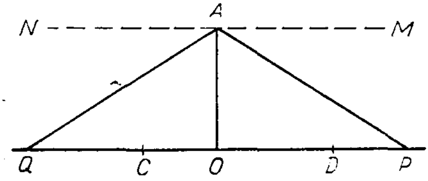


Рис. 14.

Евклид не утверждает, что это предложение очевидно, он лишь говорит, что он не может свести его к более простому или более правдоподобному утверждению. В течение столетий люди пытались сделать то, чего не смог сделать Евклид: свести это предположение к более простой идее.

Некоторое упрощение было достигнуто. Нетрудно показать, что если действительно через точку A можно провести одну и только одну прямую линию, параллельную данной прямой, то заключение о том, что сумма внутренних односторонних углов равна 180°, следует автоматически. Но можем ли мы предположить, что такая параллельная прямая существует и что она единственна? Как это можно доказать? Чтобы установить параллельность двух прямых, нужно продолжить их до бесконечности и убедиться, что они нигде не пересекаются, но ведь бесконечность — это место, где никто из нас не был.

Давайте-ка рассмотрим этот вопрос поподробнее. На рис. 14 $AO \perp CD$, P и Q — две точки на прямой CD, Q лежит слева от точки O, P — справа, причем $OQ=OP$. Проводим направо из точки A луч AM так, чтобы угол OAM был равен 90°. Луч AN проведен налево от точки A, и угол OAN тоже равен 90°.

Так как сумма двух прямых углов равна 180°, то NAM — прямая линия.

Согласно Евклиду, при перемещении точки P вправо вдоль CD прямая AP будет все более приближаться к прямой AM, но никогда с ней не сольется. То же самое можно сказать относительно прямой AQ: прямая AQ, неограниченно приближаясь к AN, никогда с ней не сольется.

Возможно, Евклид и прав. Давайте назовем это *Возможностью 1*. Она соответствует Вселенной 1.

А что случилось бы, если он неправ? Евклид утверждает, что AP может подходить как угодно близко к прямой AM, но нигде с ней не сольется, для любого конечного расстояния OP. Но здесь мыслимы еще две возможности.

Возможность 2. Для некоторого конечного отрезка OP прямая AP может в действительности совместиться с прямой AM.

Возможность 3. С другой стороны, может оказаться, что прямой AP не удастся подойти как угодно близко к прямой AM (рис. 15). Это значит, что при движении точки P вправо луч AP

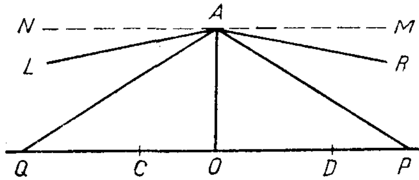


Рис. 15.

приближается к некоторому лучу AR, не сливаясь с ним, причем AR лежит ниже AM. Точно так же при движении точки Q влево AQ приближается к лучу AL, лежащему ниже AM, не сливаясь с ним.

Если вы сделаете чертеж, иллюстрирующий Возможность 2, или взгляните на тот, который я нарисовал для Возможности 3, вы скажете, что оба эти чертежа выглядят в равной степени неправдоподобно.

Я согласен с вами: для людей, воспитанных подобно нас с вами, они действительно выглядят неправдоподобными. Веками математики считали, что такие предположения невозможны, хотя их логическую несостоятельность никому не удавалось доказать. Примерно в 1830 году два математика — Лобачёвский в России и Бойаи в Венгрии,— работавшие независимо друг от друга, опубликовали статьи, где допускали Возможность 3 как вполне разумную точку зрения. В 1854 году Риман показал осуществимость Возможности 2.

Не удивительно, что Возможность 3 заметили раньше, чем Возможность 2, потому что последняя имеет весьма странное следствие. Согласно Возможности 2 точка P может пройти некоторое конечное расстояние, скажем k, вправо от точки O, и при этом луч OP сольется с AM. Иными словами, прямая NM пересечет прямую CD на расстоянии k правее точки O. Но рассматривая точку Q, мы таким же образом приходим к заключению, что прямая NM пересечет прямую CD на расстоянии k левее точки O. Однако две прямые могут пересечься лишь в одной точке. Поэтому точка, в которую вы попадаете, пройдя расстояние k вправо от O, должна совпадать с точкой, лежащей на расстоянии k влево от O. Другими словами, прямая линия должна вести себя как замкнутая кривая (окружность).

Но почему бы и нет?

Если допустить Возможность 2, то оказывается, что параллельных линий вообще не существует. Любая прямая, проведенная через точку A, неизбежно пересечет прямую CD.

Возможность 3 не так решительна. Если угол RAM очень мал, скажем составляет 1/1 000 000 долю градуса, то мы едва ли сможем отличить эту геометрию от евклидовой. Зато окажется, что через точку A можно провести бесконечное количество прямых, параллельных CD; для этой цели подходят все прямые, составляющие с AM угол, меньший 1/1 000 000 доли градуса. Но без крайне точных измерительных приборов было бы очень трудно отличить этот узкий пучок параллельных прямых от единственной параллельной прямой Евклида.

СФЕРИЧЕСКАЯ ГЕОМЕТРИЯ

Рассмотрим жизнь существ, чьей вселенной является поверхность шара. Эта вселенная не является точной иллюстрацией к геометрии Римана — Вселенной 2, но она проливает на нее некоторый свет и поясняет несколько логических положений. К тому же шар так хорошо знаком, что легко понять все рассуждения. При чтении этого раздела хорошо иметь под рукой глобус или мяч, а также кусок пластилина, глины или что-нибудь в этом роде.

Прежде всего поверхность шара удовлетворяет предположениям 1) и 2) (см. стр. 59). Твердое тело может свободно перемещаться по его поверхности. Если вы, например, придадите пластилину форму Индии и приложите его к глобусу, вы заметите, что он может свободно скользить по его поверхности. Вы можете продвинуть его так, чтобы он покрывал Европу, вы можете свободно перемещать его вокруг глобуса и все это — не отрывая пластилина от поверхности шара. Предположение 3) состоит в том, что свойства пространства одинаковы во всех точках — это тоже верно для шара: все точки сферической поверхности идентичны.

Если пластилин затвердеет, у вас получится твердое тело. Двигая его по поверхности шара, вы можете сравнивать длины в различных местах шара. Можно вернуться к предыдущим рисункам (с транспортиром или с линейкой и подковой) и убедиться, что все рассуждения справедливы и для поверхности шара.

Мы можем говорить о длинах и углах на сферической поверхности: длину можно измерить с помощью нитки, натянутой по поверхности, а в качестве прямого угла взять угол, под которым Гринвичский меридиан пересекает экватор.

Что означает понятие «прямая линия» для такой вселенной? Это кратчайшее расстояние между двумя точками. Как подчеркивалось ранее, вселенная — это теперь поверхность шара, поэтому кратчайшее расстояние означает кратчайший путь по поверхности, никаких тоннелей делать нельзя. (Нам, смотрящим на шар со стороны, такой путь кажется кривым.) Экватор, например, является кратчайшим путем между двумя лежащими на нем точками; точно так же, чтобы попасть кратчайшим путем из Лондона на Золотой Берег, нужно спуститься прямо по Гринвичскому меридиану. С точки зрения земной поверхности это «прямые линии»: будучи кратчайшими расстояниями между двумя точками, они соответствуют определению.

Недостатком этой модели с точки зрения геометрии 2 является то, что «прямые линии» пересекаются в двух точках; например, любые два меридиана пересекаются на Северном и на Южном полюсах. Но, поскольку мы находимся на некотором ограниченном участке земного шара, это нас не должно беспокоить.

Сферическая вселенная прекрасно иллюстрирует одно из положений геометрии 2 — параллельных прямых не существует. Любые

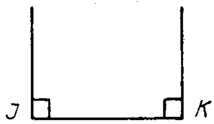


Рис. 16.

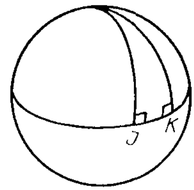


Рис. 17.

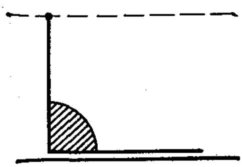


Рис. 18.

две «прямые линии» пересекаются. В геометрии Евклида, если вы с двух концов отрезка прямой восставите перпендикуляры, вы получите пару параллельных прямых (рис. 16). Но попытайтесь сделать то же самое на поверхности шара. Возьмите точки *J* и *K*, например, на экваторе (рис. 17). Две «прямые», проведенные под прямыми углами к *JK*, будут меридианами, которые пересекутся на Северном полюсе. Итак, на поверхности шара параллельных прямых не существует.

Интересным возражением на это является вопрос: «А как же железнодорожные пути?». Ведь рельсы идут на некотором постоянном расстоянии друг от друга, оставляя место для колесной оси. Как же они могут пересечься?

Переведем это на язык геометрии. Если положен один рельс, то, приладив перпендикулярно к нему скользящий стержень, мы можем начертить на Земле путь второго рельса (рис. 18). Согласно евклидовой геометрии, линия, описываемая точкой, движущейся на некотором постоянном расстоянии (измеренном по перпендикуляру) от данной прямой, есть параллельная прямая. Что же происходит с этой конструкцией на поверхности шара?

Пусть наш первый рельс идет вдоль экватора, тогда второй рельс должен идти на постоянном расстоянии от экватора. Любая из параллелей широты может служить в качестве второго рельса. Можно, например, взять полярный круг. Можно себе представить гигантский экипаж, который двигался бы по такой гигантской колее: одной парой колес по экватору, а другой — по полярному кругу. И, конечно, экватор и полярный круг не пересекаются. Значит, они параллельны? Да, но они не являются параллельными прямыми. Полярный круг не является прямой линией, поскольку он не является кратчайшим расстоянием между двумя точками на нем. Если нужно попасть из одной точки полярного круга в противоположную, то короче будет пройти через Северный полюс.

Не думайте, что геометрия на поверхности шара целиком и полностью отличается от евклидовой. Большинство евклидовых теорем, которые не зависят от идеи параллельных прямых, остаются верными для сферической геометрии. Для нее сохраняются, например, и теория конгруэнтных треугольников, и свойство равенства углов при основании равнобедренного треугольника. Так же как и в евклидовой геометрии, перпендикуляр, восставленный из середины отрезка AB, является геометрическим местом точек равноотстоящих от концов отрезка A и B.

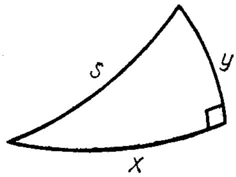


Рис. 19.

Две теоремы, базирующиеся на идее параллельных прямых, являются неверными для поверхности шара: это теорема о том, что сумма углов треугольника равна 180°, и теорема Пифагора.

Легко привести пример, показывающий, что сумма углов треугольника на сфере не равна 180°. Рассмотрим треугольник, построенный следующим образом. Из вершины, расположенной на Северном полюсе, опускаемся по Гринвичскому меридиану до экватора; дальше поворачиваем на восток и проходим четверть экватора; теперь мы находимся на долготе 90°. Затем идем вдоль меридиана к Северному полюсу. В этом треугольнике все три угла прямые. Поэтому сумма углов равна 270°. Отсюда видно, что на поверхности шара сумма углов треугольника не является фиксированной величиной. Чем больше площадь треугольника, тем больше сумма его углов.

На этом же треугольнике ясно видно, что и теорема Пифагора неверна для сферы: треугольник прямоугольный, между тем все три стороны его равны (рис. 19). Если принять радиус земного шара за единицу, то формула, соответствующая теореме Пифагора, будет иметь вид:

$$\cos s = \cos x \cdot \cos y.$$

Буквы x, y и s здесь имеют те же самые значения, что и в формулах, приведенных ранее в этой главе (в разделе «Число измерений»): x и y — расстояния вдоль двух «прямых» на поверхности шара, расположенных под прямым углом друг к другу, а s — длина третьей стороны треугольника, гипотенузы.

Геометрия сферы очевидным образом отличается от геометрии плоскости. Возникает вопрос: могли бы люди прожить на такой сферической вселенной тысячи лет и считать, что они живут на плоскости, пока не пришел бы какой-нибудь Эйнштейн? Вполне возможно принять сферу за плоскость; вековой спор относительно формы поверхности Земли, предположения, что Земля — плоскость,— ясное тому свидетельство. Но ошибка возможна только до тех пор, пока люди живут на очень малом участке сферической поверхности. Очень малая часть сферической поверхности почти не отличима от столь же малого участка плоскости. Муравью, живущему в пустыне Сахара, простительно верить, что Земля плоская.

Выше упоминалось, что сумма углов треугольника на сфере зависит от его площади. Для очень малых треугольников площадь почти равна нулю и сумма углов $\sim 180°$.

Для таких крошечных сферических треугольников верна и обычная теорема Пифагора. Это кажется странным, потому что приведенная выше формула с косинусами на вид очень отличается от формулы $s^2=x^2+y^2$. Но в действительности для очень малых x, y и s обе формулы прекрасно согласуются, $\cos x$ можно представить в виде ряда:

$$\cos x = 1 - \frac{1}{2}x^2 + \frac{1}{24}x^4 - \frac{1}{720}x^6 + \ldots$$

(см. любой учебник по дифференциальному исчислению). Этот ряд, между прочим, дает нам возможность объяснить нашему другу-ангелу термин «косинус», поскольку обычное школьное объяснение косинуса, использующее чертеж, для этой цели не подходит.

Если x мало, то величина x^4 и более высокие степени x малы по сравнению с x^2. Соответственно этому для малых $\cos x$ можно представить с высокой степенью точности просто как $1-\frac{1}{2}x^2$; более высокие степени x не оказывают влияния на сумму ряда. Например, если $x=0{,}001$, то выражение $1-\frac{1}{2}x^2$ дает значение $\cos x$ с точностью до 12 десятичных разрядов, чего, разумеется, совершенно достаточно. Используя аналогичные приближения для $\cos y$ и $\cos s$, мы получим, что уравнение $\cos s = \cos x\cdot\cos y$ принимает вид:

$$1-\frac{1}{2}s^2=\left(1-\frac{1}{2}x^2\right)\left(1-\frac{1}{2}y^2\right).$$

Перемножая скобки в правой части уравнения и вычитая 1 из обеих частей, получим:

$$\frac{1}{2}s^2=\frac{1}{2}x^2+\frac{1}{2}y^2-\frac{1}{4}x^2y^2.$$

Но так как x и y малы, то произведение x^2y^2 скажется только в 12-м десятичном разряде или что-нибудь около этого, поэтому мы можем им пренебречь. Умножая то, что остается, на 2, получим

$$s^2=x^2+y^2,$$

т. е. знакомую нам теорему Пифагора.

Мы были подготовлены к этому результату, так как уже заметили, что малые участки сферической поверхности похожи на малые участки плоскости. Вычисления лишь подтвердили то, что мы уже знали. И, однако, вычисления обратили наше внимание на некоторые факты, относящиеся к косинусам, которые мы, возможно, никогда бы не заметили, не проделав эти вычисления. Это один из путей математических открытий. Из опыта мы знаем некоторую истину. Есть у нас и формулы, в которые облечена эта истина. Пытаясь перевести наши интуитивные знания в формальное доказательство, мы открываем новые аспекты знакомых формул.

ПРОСТРАНСТВО КОНЕЧНОЕ, НО БЕЗГРАНИЧНОЕ

Вселенная бесконечна — такова традиционная точка зрения. Предположим, вы начали путешествие в каком-либо направлении и двигаетесь по прямой линии. Что помешает вам продвигаться сколь угодно далеко? Конечно же, мы не найдем, что Вселенная окружена мраморной стеной, через которую невозможно пройти. Но если это так, то мы можем идти по прямой сколько нам угодно; иными словами, начиная от Земли, Вселенная простирается бесконечно во всех направлениях. Если же это не так, то должен существовать какой-то барьер, окружающий нас. А что находится по ту сторону барьера? Разве не снова пространство? Даже если вы допускаете существование барьера, разве вы не должны признать, что по ту сторону барьера то же что-то есть? В самом деле, пространство все равно бесконечно, даже если тут и там разбросаны барьеры.

Некоторые утверждают, что теория относительности — очень сложная теория. Но разве старая теория была простой? Даже сама идея существования бесконечной вселенной — и та была потрясающей. Некоторые философы придерживались точки зрения, что бесконечность существует только как возможность. Безгранично количество слов, которое я *мог бы* произнести, но количество слов, которое я *в самом деле* произношу, к счастью, конечно. Существует ли противоречие в представлении о действительно бесконечной вселенной? Я могу только сказать, что я этого не знаю.

С появлением теории относительности старый спор получил новое толкование. Раньше утверждалось: либо вы можете двигаться вечно вперед, либо вы натолкнетесь на барьер. Но математики открыли третью возможность. Может существовать вселенная конечных размеров — занимающая конечный объем и все же не имеющая барьеров. Именно эта идея была одно время очень популярным предметом споров: «пространство конечно, но безгранично».

Вселенная на поверхности шара, описанная в предыдущем разделе, является как раз примером такой возможности. На земной поверхности нет никаких барьеров — если вы пойдете по «прямой линии» на Земле, вы можете продвигаться сколько вам захочется (я допускаю, что у вас есть средства для преодоления морей, гор, ледников и т. д.). Между тем поверхность Земли конечна — она составляет примерно 510 млн. *км*^{2}. Вот вам и иллюстрация третьей возможности: если вы достаточно долго пойдете по «прямой линии», вы можете вернуться к началу своего путешествия!

Возможно, вы считаете подобное предположение абсурдным. Но как вы докажете, что оно физически невозможно? Прямую линию можно представить как конкретное физическое тело, например как нитку, туго натянутую между двумя точками. Представим себе, что у вас с приятелем имеется неограниченный запас веревки. Вы протягиваете ее, скажем, от Солнца до звезды Сириус, затем, выпуская понемногу веревку, вы начинаете пятиться друг от друга, следя за тем, чтобы веревка была туго натянута и проходила вблизи Сириуса и Солнца. Ваша веревка должна представить прямую линию, соединяющую эти две звезды. Откуда вы знаете, что рано или поздно не столкнетесь друг с другом спинами, точно так же, как если бы вы двигались по экватору, пятясь друг от друга. Если вы скажете: «Нет, это невозможно», вы встанете на ненаучную точку зрения: вам предлагается физический эксперимент, вы же заявляете, что можете с уверенностью предсказать результат, не проделывая эксперимента (ведь «невозможно» выражает полную уверенность). С бо́льшим правом можно было бы сказать: «Если Вселенная такова, какой я себе ее представляю, то ваша идея невозможна», но это ведь то же самое, что заявить: «Если эта Вселенная подчиняется законам евклидовой геометрии, ваше предположение невозможно»; но мы уже видели, что существуют и другие геометрии, и Вселенная может быть построена по законам одной из них. Несомненно, Вселенная всегда таит какой-нибудь сюрприз. Люди часто думали, что они уже все знают о мире, но всегда оказывалось, что они ошибаются.

ТРЕХМЕРНАЯ СФЕРИЧЕСКАЯ ВСЕЛЕННАЯ

Поверхность шара имеет только два измерения. Двух чисел, например широты и долготы, достаточно, чтобы определить ваше местоположение. В идее о том, что Вселенная имеет форму поверхности шара, могло бы воплотиться представление об общей теории относительности у людей, долгое время считавших, что они живут на евклидовой плоскости. Заметьте, что идея сферы нам понятна, но она не была бы так понятна существам, живущим в сферической вселенной. Мы легко представляем себе сферу, поскольку мы живем в трехмерном пространстве. Но люди, которые действительно жили бы на поверхности шара, знали бы только два измерения: третье же измерение и находящийся в нем шар — это вещи, до которых они могли бы дойти только абстрактным математическим рассуждением, но не смогли бы наглядно себе его представить или осознать своими чувствами.

Точно так же, если мы хотим сконструировать модель «конечной, но безграничной» трехмерной вселенной, мы должны быть готовы мыслить в терминах четырех измерений и уметь представить себе тот образ, который в четырехмерном пространстве соответствует шару. С точки зрения математики никаких трудностей не возникает.

Быть может, полезно будет начать с более простых вещей. На плоскости, которая имеет два измерения, каждая точка определяется парой чисел x и y. Точки плоскости, находящиеся на единичном расстоянии от начала, образуют окружность. Из теоремы Пифагора мы знаем, что для таких точек справедливо соотношение $x^2+y^2=1$. Мы можем представить себе вселенную, состоящую только из точек этой окружности. Это была бы одномерная вселенная. Если бы в такой вселенной мы продолжали идти в одном направлении, то вернулись бы в точку, откуда начали свой путь.

В пространстве трех измерений каждая точка определяется тремя числами x, y, z. Точки, расположенные на единичном расстоянии от начала, образуют сферу. Она задается уравнением $x^2+y^2+z^2=1$ и представляет собой модель двумерной вселенной.

До сих пор мы могли бы себе ясно представить, о чем идет речь. Но для ангела на другом конце провода все это были абстрактные упражнения ума. Он готов сделать следующий шаг и не понимает, почему нам это трудно.

В четырехмерном пространстве каждая точка определяется четырьмя числами x, y, z, t. Расстояние s любой точки от начала задается выражением $s^2=x^2+y^2+z^2+t^2$. Точки, находящиеся на единичном расстоянии от начала, удовлетворяют условию $x^2+y^2+z^2+t^2=1$. Они образуют гиперсферу, представляющую модель трехмерной вселенной.

Мне кажется, этих предварительных данных достаточно, чтобы начать математическую разработку вопроса; теперь можно определить свойства такой вселенной. Как уже подчеркивалось ранее, для людей трехмерной вселенной реальны только точки гиперсферы. Тот факт, что они не могут выйти за ее пределы, нисколько их не беспокоит, да они вообще не могут вообразить себе подобное. В их вселенной есть место для Земли и Солнечной системы вроде нашей, для направлений «восток», «север» и «вверх», которые все находятся в самой гиперсфере. «Выйти из гиперсферы» означает для них путешествие под прямым углом к направлениям «восток», «север» и «вверх» (одновременно), а такая идея никогда не приходит им в голову.

Я не собираюсь вдаваться в математические расчеты, а хочу лишь по аналогии со сферой вывести два свойства этой вселенной.

Вернемся ко вселенной, представляющей собой поверхность шара. Свет в природе распространяется по «прямой линии». Мы уже видели, что на поверхности шара экватор и меридиан являются «прямыми линиями». Представим себе человека, стоящего на Северном полюсе и смотрящего на весь мир. Что он видит? Рядом с собой он, конечно, увидит снег, лед и белых медведей — это его непосредственное окружение. Но взгляд его может проникнуть за пределы горизонта, потому что свет следует по контурам шара; мы бы сказали, что свет изгибается вдоль поверхности шара. Но человек этого не говорит. Если бы свет распространялся прямо в нашем обычном смысле слова, он уходил бы из пределов вселенной. Свет распространяется настолько прямо, насколько это возможно, чтобы не покидать пределов вселенной. Предположим, человек стоит так, что он смотрит прямо вдоль Гринвичского меридиана. Он сможет увидеть Гринвич, далее часть Франции, Испании и Африки (при условии, что одно не заслоняет другое); далее, он увидит большое водное пространство, льды Антарктики и Южный полюс. Но линия его

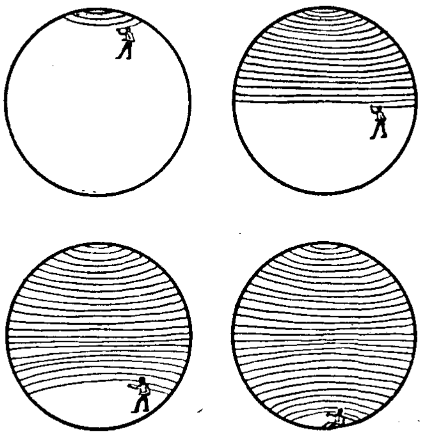


Рис. 20.

взгляда, все еще следуя по поверхности шара, пойдет вверх по меридиану, расположенному под углом 180° к Гринвичскому. За Южным полюсом он увидит Новую Зеландию, несколько островов в Тихом океане, крайнюю точку Сибири и позади всего этого — свой собственный затылок! И в каком бы направлении он ни смотрел, последнее, что он увидит, будет его собственный затылок.

Для трехмерной аналогии справедливо то же самое. Если погода была бы ясной, а наши телескопы достаточно мощными, то, в каком бы направлении мы ни смотрели, мы увидели бы самих себя, или, более точно, принимая во внимание время, которое свет затрачивает на то, чтобы обогнуть вселенную, мы увидели бы то, что было на Земле несколько миллионов лет назад.

Это первое свойство, которое мне хотелось объяснить. Теперь о втором. Как мог бы наш человек, живущий на поверхности сферы, узнать, что он живет не в бесконечном пространстве? Предположим, что погода настолько плоха, что он видит не далее, чем на километр, так что он не может ответить на вопрос только при помощи зрения. Предположим также, что у него имеется сколько угодно изоляционной ленты (или лейкопластыря). Он вбивает столб на Северном полюсе и начинает его обматывать, делая столб все шире и шире (рис. 20). Он продолжает наматывать ленту и со временем покрывает все северное полушарие и оказывается на экваторе. Но, пересекая экватор, он не замечает какой-либо резкой перемены. Окружность Земли по экватору равна 40 000 *км*, он и не замечает, что его круги постепенно становятся все меньше. Он все еще думает, что находится с внешней стороны гигантского круга с центром на Северном полюсе, и продолжает накладывать ленту. Однако по мере приближения к Южному полюсу он начинает замечать перемену. Он уже больше не находится снаружи круга, он внутри него. Чем больше ленты он наматывает, тем меньше становится его тюрьма, и, наконец, он покрывает всю Землю изоляционной лентой и запирает себя на Южном полюсе.

Давайте проведем подобную аналогию для гиперсферы. Вместо того чтобы обматывать лентой столб (делая его все толще и толще), мы можем представить себе, что начинаем с крикетного шара и увеличиваем его, накладывая краску слой за слоем. Краска у нас имеется в неограниченном количестве, и мы накладываем ее все больше и больше. По мере того как радиус шара растет, его поверхность кажется нам все менее выпуклой. Но вот приходит время, когда мы пересекаем экватор нашей гиперсферы. Поверхность разросшегося шара в это время кажется практически плоской; однако мы не замечаем никакой резкой перемены. Но, продолжая, мы через некоторое время обнаруживаем нечто странное — поверхность не просто становится плоской, она начинает выгибаться в противоположную сторону. Мы больше не снаружи окрашенной поверхности, мы — в полой сфере. По мере того как мы продолжаем красить, сфера становится все меньше и меньше, и мы прекращаем работу, когда нам уже негде двигаться. Вот что бы мы чувствовали, если бы жили в пространстве «конечном, но безграничном». И некоторые люди серьезно предполагали, что мы действительно живем в таком пространстве.

Экспериментальное доказательство — дело физиков. Математиков касается только логическая возможность этой теории. Они считают, что так могло бы быть.

ЭКВИВАЛЕНТНЫЕ ВСЕЛЕННЫЕ

Возможно, вас не удовлетворяет «изогнутая» вселенная, требующая дополнительного измерения. Ибо если вселенная представляет собой сферу (или гиперсферу), то что находится внутри нее, а что снаружи? Но те же самые возражения возникают и по отношению к «плоской» (т. е. евклидовой) вселенной. Если двумерная вселенная представляет собой плоскость, то что находится над плоскостью и что под ней?

В самом деле, математически разницы между двумя этими идеями нет. Вообразим, например, людей, мир которых — это дуга AB окружности (рис. 21). Пусть из некоторой точки P идет поток света, тогда тени от предметов на дуге AB падают на прямую CD. Представим себе теперь вселенную, построенную на CD, предметы которой *ведут себя точно так же, как тени от предметов на AB*. Эти две вселенные будут подчиняться одним и тем же законам. Существа, живущие на AB, не смогут сказать, что они не живут на CD. В действительности мы имеем здесь два разных способа описания одной и той же вселенной. Если мы описываем вселенную как часть круга, мы можем сказать, что твердые тела сохраняют постоянную длину при перемещении. Если же мы описываем вселенную как часть прямой CD, мы должны будем ввести какой-нибудь фактор (что-нибудь вроде температуры), чтобы объяснить, почему

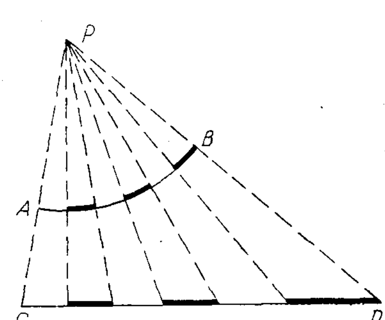


Рис. 21.

длины тел изменяются при их движении вдоль прямой. Но ни один эксперимент не позволит жителям решить, в какой вселенной они обитают. Эти две теории отличаются друг от друга не по существу, а лишь живописными подробностями. И уже дело вкуса, какой картине оказать предпочтение.

Для иллюстрации высказанной идеи[1] я приведу «плоскую» модель двумерной вселенной, реализующую Возможность 3. Такую вселенную можно было бы получить, предполагая, что предметы скользят по подходящей кривой поверхности. Автором этой модели является Пуанкаре (1854—1912).

ВСЕЛЕННАЯ ПУАНКАРЕ

Эта вселенная заключена внутри круга. В центре круга температура довольно высока, но по мере удаления от центра температура падает, достигая абсолютного нуля на границе круга. Закон изменения температуры очень прост: если a — радиус круга, то на расстоянии r от центра круга температура будет равна $T=a^2-r^2$.

Если какой-нибудь предмет перемещается в этой вселенной, его величина зависит от изменений температуры. Предположим, что длина предмета пропорциональна температуре T. На границе круга, где $r=a$ и $T=0$, длина предмета сократится до нуля. Ширина предмета изменяется точно так же, как его длина.

Но обитатели этой вселенной ничего не знают о температуре. Мы предполагаем, что у них нет нервов, чувствительных к теплу (или холоду), и они не могут ощущать температуру непосредственно. Они также не могут измерять температуру с помощью термометров. В обычном термометре используется свойство ртути расширяться в большей степени, чем стекло. Но в этой вселенной все тела расширяются и сжимаются одинаково. Если существо имеет в длину 180 *см*, находясь в центре своей вселенной, то, перейдя в более холодную часть, оно также будет 180 *см* длины. Нам кажется, что его длина уменьшилась в два раза, но ведь и длина 30-сантиметровой линейки уменьшилась в такой же степени, и поэтому наше существо всегда будет оставаться в шесть раз длиннее линейки.

На это, конечно, можно возразить, что существо заметит при достижении границы вселенной, что длины тел становятся равными нулю. Но дело в том, что никто не может достичь границ вселенной (рис. 22), потому что, как только существо начинает движение по направлению к этой границе, его размеры становятся все меньше, и чем ближе оно подходит к границе, тем быстрее сокращаются его размеры. Ему самому кажется, что оно делает шаги равной длины, а мы видим, что шаги становятся все короче и короче, и закон таков, что сколько бы шагов оно ни сделало, оно никогда не достигнет

[1] Т. е. идеи о том, что возможны различные модели одной и той же (с точки зрения математики) вселенной.— *Прим. ред.*

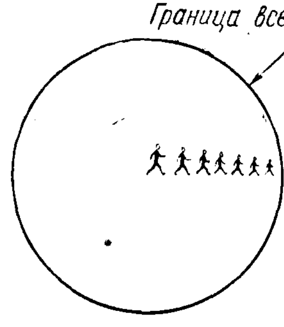


Рис. 22.

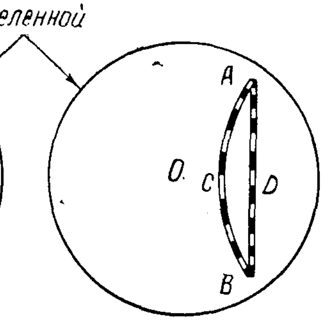


Рис. 23.

границ вселенной [1]. Граница, с точки зрения этого существа, находится на бесконечном расстоянии от центра. Линии, пересекающиеся на границе, параллельны. Таким образом, хотя нам кажется, что эта вселенная занимает конечное пространство, для существ, населяющих ее, она бесконечна.

Что представляет собой «прямая линия» в такой вселенной? По определению, «прямая линия» есть кратчайший путь между двумя точками. Заметим, что длина пути — это длина, измеренная самими созданиями, населяющими вселенную. ADB на рис. 23 — линия, которую бы мы назвали прямой. Темные и светлые участки на ADB обозначают 15 звеньев цепи, соединяющей точки A и B. Звенья этой цепи, с точки зрения жителей вселенной, равны по длине. Звенья, находящиеся рядом с точкой D, лежат, конечно, ближе к O, центру вселенной. Если бы эти звенья поместить в более холодные участки, около точек A или B, то они приняли бы такие же размеры, какие имеют сейчас звенья, прилегающие к A и B.

Другая цепь звеньев образует путь ACB. Звенья этой цепи имеют те же размеры, что и звенья цепи ADB. Но поскольку точка C расположена ближе к центру, чем точка D, то цепь ACB содержит меньше звеньев, чем цепь ADB. Путь ACB «короче», чем путь ADB. Для обитателей этой вселенной ACB есть «прямая линия», соединяющая точки A и B; для цепи ACB нужно меньшее количество звеньев, чем для любого другого пути.

---

[1] Это можно показать с помощью несложных вычислений. Отрезок ds в центре круга, перенесенный на расстояние r от центра, будет иметь в обычном смысле длину $dr = \frac{a^2 - r^2}{a^2} ds$; для жителя нашей новой вселенной этот отрезок попрежнему имеет длину ds. Расстояние от центра круга до границ житель вселенной оценивает как $\int ds = \int_0^a \frac{a^2}{a^2 - r^2} dr$, что равно бесконечности.

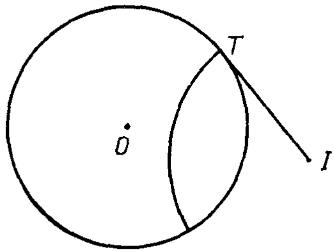


Рис. 24.

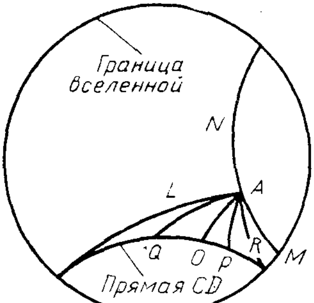


Рис. 25.

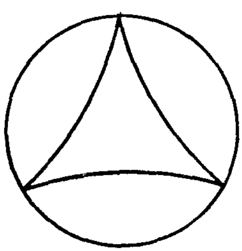


Рис. 26.

Имеется очень простой способ построения «прямой линии». «Прямые линии» являются на самом деле (т. е. у нас на чертеже) окружностями. Однако не любая окружность подойдет для этой цепи, а только такая, которая пересекает граничную окружность под прямым углом (рис. 24). Если *I* — точка, расположенная за пределами граничной окружности, а *IT* — касательная к этой окружности в точке *T*, то дуга окружности радиуса *IT* с центром в точке *I* будет «прямой линией». Через две точки вселенной *A* и *B* может быть проведена одна и только одна «прямая линия». Такую форму, естественно, будет иметь натянутая проволока, луч света пойдет по этому же пути. Так, на рис. 23, где изображены две цепи, *ACB* имеет как раз ту форму, какую приняла бы цепь, натянутая между точками *A* и *B*. Обитатели вселенной могли бы проверить это визуально; поскольку свет следует по пути *ACB*, то для обитателя, смотрящего из точки *B*, точка *C* заслонила бы точку *A*, а это как раз есть способ убедиться, что три точки *A*, *B*, *C* лежат на прямой линии. Так, кратчайший путь между двумя точками будет выглядеть прямым для обитателей этой вселенной, хотя нам он прямым не кажется.

Ранее, рассматривая чертеж, иллюстрирующий Возможность 3, я отмечал, что нам он не кажется правильным. На рис. 25 вы видите этот же чертеж в том виде, какой он имел бы во вселенной Пуанкаре: *AO* — «прямая», перпендикулярная «прямой» *CD*, точка *P* расположена справа от *O*, точка *Q* — слева. «Прямая» *NAM* — перпендикуляр к *OA*. При движении *P* к «бесконечности», т. е. к границе вселенной, «прямая» *AP* приближается к положению *AR*. Между *AR* и *AM* остается значительный угол. Если мы проведем «прямую» из точки *A* так, чтобы она лежала внутри углов *LAN* или *RAM*, то получим «прямую», которая не пересекается с CD. Можно провести бесконечное множество таких прямых.

Если бы мы взяли небольшой участок этой вселенной, температура в его пределах менялась бы очень незначительно. Соответственно геометрия этого маленького участка не отличалась бы от евклидовой. Обитатели этого участка, если бы их возможности путешествовать были ограничены, могли бы думать, что они живут на евклидовой плоскости. В этом есть мораль и для нас, поскольку мы живем в пределах Солнечной системы, которая является крошечным пятнышком по сравнению с расстояниями между звездами.

Во вселенной Пуанкаре сумма углов очень малого треугольника почти равна 180°, но, чем больше треугольник, тем меньше сумма его углов. Взгляните на треугольник, образованный «прямыми» AL, AR и CD (рис. 25). Углы, прилегающие к границам вселенной, равны нулю. Угол AR значительно меньше 180°, так что сумма углов треугольника меньше 180°. В треугольнике, изображенном на рисунке 26, сумма углов равна нулю.

Измеряя углы большого треугольника, можно было бы выяснить, какая из трех возможностей реализуется в нашей Вселенной.

Глава седьмая

АЛГЕБРА БЕЗ АРИФМЕТИКИ

Пусть У — университет, Г — греческий, П — профессор, тогда ГП — профессор греческого.

Льюис Кэрролл

В одном старом определении математики говорится, что это наука о числах (арифметика) и фигурах (геометрия). Числа и фигуры — очень разные вещи, почему же их объединяют в один предмет? Безусловно верно, что фигуры могут быть измерены, и с тех пор, как Декарт ввел в употребление систему координат, стало возможным перевести любую геометрическую задачу на язык алгебры. Но представление о математике как об арифметике плюс геометрии появилось гораздо раньше Декарта (1596—1650). В традиционную геометрию, восходящую к древним грекам, арифметика входила едва ли не вся.

По-видимому, арифметика и геометрия (по исторической случайности) оказались первыми науками, которые получили полностью логическую форму. В дальнейшем и многие другие предметы также трактовались математически.

Если сравнить математику, биологию и искусствоведение, можно найти ряд математических черт, присущих всем этим наукам. Искусствоведение, насколько я знаю, не претендует на звание дедуктивной науки. Оно не начинает с четкого определения художественной ценности произведения (я не говорю «красоты», потому что многие великие произведения искусства не красивы) и не оценивает картины, симфонии или книги на основе этого определения. Если бы мы достаточно хорошо знали самих себя, если бы мы с большой точностью могли предсказывать нашу реакцию на ту или другую вещь, то была бы возможна математическая теория искусства. Я могу себе представить антрополога, определяющего точную пропорцию эстетических качеств, привлекательную для примитивного человека, или составляющего оптимальную смесь секса, садизма и сентиментальности, требующуюся современному магнату кинопромышленности. Но вообще искусствоведение не является упражнением в дедукции, да это бы ему ничего и не дало.

Биология отчасти дедуктивна. Признаются некоторые общие положения — вроде того, что каждому животному нужен какой-нибудь источник энергии для поддержания жизни,— и на их основе можно начинать строить систему, рассматривающую, какого рода организмы могут жить в той или иной среде. Но в биологии многие термины неопределяемы, в частности термин «жить»; наряду с логическими выводами в ней всегда требуются наблюдения, опыт и здравый смысл. Это не строго формальная дисциплина.

Математические науки стремятся к полностью дедуктивному подходу. Предполагается, что идеи настолько четко определены, что можно развивать их чисто логическим путем.

Науки вообще имеют тенденцию становиться все более математическими, так как с развитием какой-нибудь науки ученые постепенно начинают понимать, о чем они говорят, и все больше отдают себе отчет в методах, при помощи которых они делают свои заключения. Поскольку идеи становятся все более ясными, возрастает возможность их логического развития. Однако должны быть и некоторые границы этого процесса. Если бы, например, существовала математическая теория истории, дающая возможность предсказать будущее, то полученные таким образом знания заставили бы людей действовать иначе [1]. Точно сформулированная теория оценки художественных произведений произвела бы аналогичное действие: она заставила бы людей изменить свои вкусы.

Итак, математика оперирует четко определенными понятиями или идеями. Нет причин для того, чтобы математические символы употреблялись только для чисел (как в арифметике, алгебре, тригонометрии) или для точек (как в геометрии). Они могут применяться для всего, чего угодно. Что бы они ни обозначали, мы развиваем теорию в соответствии со свойствами, которыми обладает сам предмет.

АЛГЕБРА ЛОГИКИ

В качестве примера введем символы для слов «и» и «или». Поскольку смысл этих слов хорошо понятен, наши идеи будут иметь необходимые точность и ясность [2]. Введем следующие символы:

$+$ для обозначения союза *и*,

$\cdot$ для союза *или* [3],

a для высказывания «Альфред говорит правду»,

---

[1] Эта идея увлекательно развита в романе Э. Хайэмза «Астролог» (E. Hyams, «The Astrologer»).

[2] Я не знаю, обладает ли какая-нибудь идея абсолютной точностью, но для формальной теории важно лишь, чтобы идея была достаточно точна для того, что вы намерены с ней делать. [«Смысл этих слов», во всяком случае слова «или», не так уж «понятен», поскольку имеется по крайней мере *два* понимания слова «или»: 1) *неразделительное* «или» (когда высказывание «*A* или *B*» считается истинным, если *по крайней мере одно* из высказываний «*A*» или «*B*» истинно — см. ниже, стр. 79), 2) *разделительное* «или» (когда высказывание «*A* или *B*» считается истинным, если *одно и только одно* из высказываний «*A*» и «*B*» истинно — см. стр. 79). Поскольку в дальнейшем фигурируют *оба* понимания, читатель должен внимательно следить за вводимыми автором соглашениями.— *Ред.*]

[2] Эти обозначения не общеприняты; чаще обозначают «или» через $+$ (а в литературе по математической логике через $\vee$), а «и» — через $\cdot$ (соответственно через &).— *Прим. ред.*

b для высказывания «Бетти говорит правду»,

c для высказывания «Чарлз говорит правду»,

$=$ для обозначения отношения *равносильно тому, что.*

Если написано любое сочетание этих символов, мы используем приведенный выше «словарь» и можем судить, каков результат: верный, неверный или бессмысленный. Мы заменяем каждый символ словом или словами, вместо которых он стоит, и выносим приговор получившемуся предложению. Посмотрите на равенства (1), (2) и (3). Что вы думаете о них? (Я отвечу на этот вопрос ниже (после звездочек); если вы хотите справиться сами, не читайте сразу ответ.)

(1) $$a+b=b+a$$

(2) $$a\cdot b=b\cdot a$$

(3) $$a\cdot(b+c)=a\cdot b+a\cdot c$$

* * * * *

Равенство (1) гласит: «То, что Альфред и Бетти говорят правду, равносильно тому, что Бетти и Альфред говорят правду». (Если расшифровать символы дословно, получится немного длиннее: «Альфред говорит правду и Бетти говорит правду...» и т. д. Символ, однако, тот же самый. Последующие примеры я сокращаю таким же образом.) Равенство (1), таким образом, правильно, хотя истина, которую оно выражает, несколько тривиальна.

Равенство (2) гласит: «То, что Альфред или Бетти говорит правду, равносильно тому, что Бетти или Альфред говорит правду». Это опять очевидная истина.

Равенство (3)можно было бы применить к ситуации, где показания свидетелей противоречивы: Бетти и Чарлз излагают подтверждающие друг друга факты, но противоречащие тому, что рассказывает Альфред. Равенство (3) гласит: «То, что Альфред говорит правду или Бетти и Чарлз говорят правду, равносильно тому, что Альфред или Бетти говорят правду и Альфред или Чарлз говорят правду». Это утверждение тоже верно.

Замечательно, что (1), (2), (3) — это равенства, которые верны в обычной школьной алгебре, где $+$ означает «плюс», а $\cdot$ означает «умножить на». Одна и та же закономерность встречается в двух совершенно различных областях применения. Значит, привычный нам способ обращения с равенствами, связывающими числа, может быть распространен и на логические истины. Перемножая, например, $(a+b)$ и $(c+d)$ и пользуясь нашим «словарем», мы получим немного громоздкое, но верное логическое утверждение (где, скажем, d означает «Дэвид говорит правду»).

Замечательно и то, что знаки $+$ и $\cdot$ можно менять местами. Если вы захотите употребить $+$ вместо слова «или», а $\cdot$ вместо «и», то, произведя алгебраические действия, вы всегда будете получать верный логический результат. В обычной алгебре знаки + и · нельзя менять местами. Следовательно, хотя все результаты обычной алгебры верны для алгебры логики [1], не все результаты логической алгебры верны для обычной алгебры. Например, равенство

$$a+b\cdot c=(a+b)\ (a+c), \tag{4}$$

полученное из (3) заменой + на · и · на +, справедливо в алгебре логики; оно означает: «То, что Альфред говорит правду и Бетти или Чарлз говорят правду, равносильно тому, что Альфред и Бетти или Альфред и Чарлз говорят правду». Если же вместо букв поставить числа, то подобная замена окажется неправомерной.

Но, в конце концов, вовсе не обязательно, чтобы правила любой алгебры совпадали с правилами алгебры обычных чисел. Коль скоро имеются определенные правила, по которым следует производить операции над символами, не так уж важно, откуда взялись эти правила.

АЛГЕБРА КЛАССОВ

Еще одна алгебра, также относящаяся к логике,— это алгебра классов. Мы можем воспользоваться, например, символом s [2] для обозначения класса «всех малых предметов» и g [3] — для класса «всех зеленых предметов»; sg пусть означает «все малые зеленые предметы». Это определение очень близко к обычной речи. Буква s выражает свойство «быть малым», g — свойство «быть зеленым»; мы сочетаем обе буквы так же, как слова «малый» и «зеленый» в предложениях вроде «Это маленькая зеленая дверца». Рассматриваемое как алгебраический символ выражение sg имеет вид произведения; и оно, как мы увидим, действительно ведет себя как произведение.

Сумма $s+g$ определяется довольно странным образом: $s+g$ означает все вещи, являющиеся либо малыми, либо зелеными, но не обладающими одновременно обоими этими качествами [4]. Нельзя сказать, чтобы это было очень естественное определение, но оно обеспечивает простейшие правила для выполнения вычислений [5].

---

[1] Конечно, при условии, что они имеют логическую интерпретацию; я не давал никакой интерпретации, например, для выражения $a-\frac{1}{2}b$.

[2] От англ. small — маленький, небольшой.— *Прим. перев.*

[3] От англ. green — зеленый.— *Прим. перев.*

[4] Такое определение *суммы классов* соответствует операции *разделительного* «или» над *высказываниями* о *свойствах* этих классов (см. примечание редактора к стр. 77). Чаще сумма классов определяется как класс, включающий элементы *хотя бы одного* из классов-слагаемых (а не *в точности одного*, как здесь).— *Прим. ред.*

[5] Правила вычислений для операции сложения классов, упомянутой в предыдущем примечании, ничуть не «сложнее» приведенных в тексте.— *Прим. ред.*

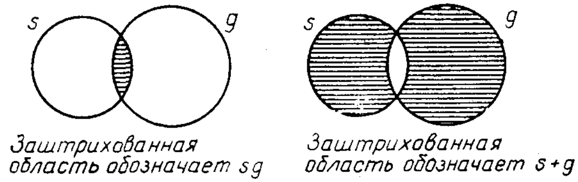


Рис. 27.

Смысл sg и $s+g$ иллюстрирует рис. 27. Предположим, что круг s охватывает все имеющиеся в мире малые предметы, а круг g — все зеленые; тогда заштрихованные области будут означать соответственно sg и $s+g$.

Отметим особо, что происходит с $a+b$, когда класс a лежит целиком внутри класса b. Рис. 28 показывает, как область a проникает все глубже в область b. (Обозначим, например, через a класс «Все люди, которым еще нет шестидесяти лет», а через b — «Все люди, родившиеся после 1900 года». На рис. 28 видно, что происходит между 1950 и 1970 годами.

Вы видите, что когда класс a является частью класса b, то $a+b$ означает «имеющий свойство b, но не свойство a»[1]. Если бы случилось, что класс a разросся до таких пределов, что слился бы с b, то заштрихованная площадь на рис. 28 исчезла бы совсем. Поэтому $a+a$ означает пустоту, ничто. Мы можем использовать знак 0 для обозначения понятия «ничто». Итак, $a+a=0$. Для обозначения понятия «всё» мы используем знак 1. На рис. 29 широкий прямоугольник содержит «всё», а маленький кружок — a.

Нетрудно заметить, что $1+a$ означает «всё, за исключением a», другими словами, «не a».

Возможно, вы считаете, что $1—a$ больше подошло бы для обозначения «всё, кроме a». Если вы предпочитаете это выражение, можете писать его, поскольку $a+a=0$, $+a=—a$. Плюс и минус означают здесь одно и то же! И все-таки лучше употреблять знак +, потому что $s+g$ означает то же, что и $g+s$, а это свойство гораздо легче запомнить со знаком +, чем со знаком —. Но, повторяю, вы могли бы поставить здесь —, если вам так больше нравится.

Безусловно, $1+1=0$. Возможно, вас удивит это. Однако несложный пример поможет вам понять, в чем будет дело. Вы знаете, что в некоторых квартирах в целях безопасности отсутствуют электрические выключатели, вместо них с потолков свисают длинные шнуры. Если света нет, вы дергаете за шнурок — и свет зажигается. Если вы дернете еще раз, свет погаснет. Дернуть за шнурок два раза все равно, что не дернуть ни разу. В этом случае $1+1=0$. Эффект,

[1] Здесь допущена некоторая неточность. Вообще говоря, не следовало бы обозначать класс и свойство одной и той же буквой a. Но большого вреда от этого не будет.

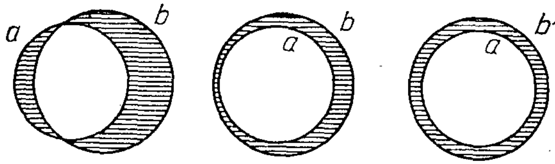


Рис. 28.

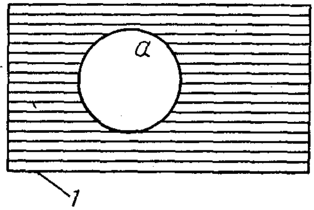


Рис. 29.

который достигается тем, что вы дергаете за шнур, в логике равносилен воздействию слова «нет». Если свет включен и вы дергаете за шнурок, то он гаснет, становится «не включен». Если же свет выключен и вы дергаете шнур, свет зажигается, становится «не выключен».

Мы видим теперь, что равенство $1+1=0$ соответствует свойству отрицания «не», $a+1$ означает «не a», $a+1+1$ означает «не не a» или просто «a». Поэтому $1+1$ должно равняться нулю. Было бы просто неверно, если бы это было не так.

Мы рассмотрели, что такое $a+a$. А что такое $a \cdot a$? Очень просто. Вернемся к нашему обозначению g для понятия «зеленый», тогда gg означает «зеленые, зеленые предметы». В поэзии сказать, что что-то было «зеленое, зеленое», значит сказать «очень зеленое», это все равно, что воскликнуть: «Ах, до чего же зелено!», или что-нибудь вроде этого. В логике это просто означает «все зеленые предметы». Поэтому $gg=g$. Аналогично будет для a: что бы ни означало a, всегда $aa=a$.

Самое замечательное, что символы, определенные таким образом, прекрасно подчиняются законам алгебры (конечно, за исключением правила $1+1=0$). Я не собираюсь приводить систематическое доказательство, а дам лишь несколько иллюстраций того, как алгебра приводит к верным результатам.

Как будет обозначаться в этой символике «a или b»? «Или» я здесь употребляю в том же смысле, что и в выражении: «Людей пропускают, если у них есть разовый входной билет или сезонный билет»; конечно, это не исключает возможности, что найдется какой-нибудь эксцентричный или щедрый господин, который, имея сезонный билет, купил также входной [1].

Взгляните на приведенные выше рисунки, и вы увидите, что «зеленый или маленький (или то и другое)» занимает обе заштрихованные части рисунков для sg и $s+g$. В действительности это есть

[1] Именно такое понимание союза «или» неявно подразумевалось на стр. 77; если же упомянутый «господин» напал бы на не менее «эксцентричного контролера», заботящегося о «простоте вычислений» (см. стр. 79), то тот не пропустил бы нашего героя или заставил бы его продать один из билетов — иными словами, «или» понималось бы им как *разделительное* (см. стр. 77).— *Прим. ред.*

$sg+s+g$. Не знаю, заметили ли вы это, но я, как только посмотрел на $sg+s+g$, сразу же подумал: «Это очень похоже на $(s+1)(g+1)$». Поскольку при умножении мы получаем $sg+s+g+1$, то $sg+s+g=(s+1)(g+1)-1$, а так как -1 все равно, что $+1$, то это тождественно выражению $(s+1)(g+1)+1$.

Последнее выражение легко интерпретировать: $(s+1)$ означает «не маленький», $(g+1)$ — «не зеленый», произведение $(s+1)(g+1)$ означает «не маленький и не зеленый», $+1$ означает «не» или «все, кроме». Поэтому все выражение означает «то, что останется, если вы удалите все, что не мало и не зелено». Но именно это значение и должно было получиться. Все предметы делятся на 4 класса:

1) маленькие и зеленые;
2) маленькие, но не зеленые;
3) зеленые, но не маленькие;
4) не маленькие и не зеленые.

Поэтому, удалив класс (4), мы оставим только те предметы, которые являются или маленькими, или зелеными, или теми и другими вместе. Вывод не слишком поразительный, но он подтверждает, что обычные законы алгебры (вместе с совершенно необычным уравнением $+1=-1$) приводят к правильным логическим заключениям.

Другой пример. Перемножив a и $(1+a)$, получим $a(1+a)=a+aa=a+a=0$. Это правильно, потому что $a(1+a)$ означает «обладающий свойством a, а также свойством не a», а это невозможно.

Легко также проверить, что $a(b+c)=ab+ac$. Мы начнем с $a(b+c)$. На рис. 30 $(b+c)$ состоит из областей *2, 3, 5, 7*. Поэтому $a(b+c)$ состоит из областей *2* и *5*. С другой стороны, ab включает *2* и *4*, ac — *4* и *5*; $ab+ac$ состоит из таких областей, которые входят либо в ab, либо в ac, но не в обе области сразу, т. е. это будут области *2* и *5*.

Выражение $a+b+c$ также представляет некоторый интерес. Оно обозначает те предметы, которые относятся к *нечетному числу* классов a, b, c, т .е. области *1, 3, 4, 7*. Это проверяется легко. Вместе с тем получается, что выражение $a+b+c$ означает одно и то же, независимо от того, считаете ли вы его результатом сложения a с $b+c$, или b с $a+c$, или c с $a+b$.

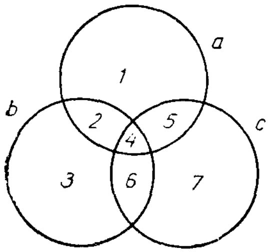


Рис. 30.

И наконец, последняя иллюстрация. Вопрос ставится следующим образом. «Если все вареные красные раки мертвы и все вареные мертвые раки красны, то значит ли это, что все красные мертвые раки вареные?» Есть много способов ответить на этот вопрос. Я рассмотрю это алгебраически.

Поскольку в нашей задаче говорится только о раках, нам не нужно вводить никакого символа для обозначения понятия «рак», мы попросту будем считать, что во всем мире нет ничего, кроме раков. Обозначим буквой b[1] слово «вареный», буквой r[2] — «красный» и буквой d[3] — «мертвый».

Как же с помощью этих символов выразить, что «все вареные красные раки мертвы»? Можно пересказать это выражение так: класс вареных красных немертвых раков является пустым. Запишем это с помощью символов: $br(1+d)=0$. Перемножив, получим $br+$ $+brd=0$, что означает $br=-brd$. Но так как минус означает то же, что плюс, то $br=brd$. Этот результат мы можем проверить по смыслу. Здесь говорится, что класс вареных красных мертвых раков — тот же самый, что класс вареных красных раков. Но так и должно быть, ибо мы знаем, что раки, обладающие свойствами «вареный» и «красный», обладают и тем свойством, что они мертвы.

Точно так же, высказывание, что все вареные мертвые раки красны, можно записать как $bd=bdr$.

В задаче спрашивается, являются ли вареными все красные мертвые раки, то есть справедливо ли уравнение $rd=rdb$.

Итак, задача формулируется следующим образом. Дано $br=brd$ и $bd=brd$; следует ли отсюда, что $rd=brd$? (Порядок букв b, r и d в произведениях brd, bdr, rdb не играет роли.)

Не видно, чтобы третье уравнение следовало из первых двух. Но все же, как проверить это точно?

Если бы третье уравнение следовало из первых двух, то это означало бы, что все значения величин b, r и d, удовлетворяющие первым двум уравнениям, удовлетворяют также и третьему. Но мы можем быстро показать, что это не так. Действительно, при $r=1$, $d=1$, $b=0$ первые два уравнения удовлетворяются, но третье обращается в $1=0$, или «все равно ничему», что неверно.

Следовательно, нельзя утверждать, что все красные мертвые раки вареные.

ГРУППА ДВИЖЕНИЙ

Оставим вопросы формальной логики и рассмотрим символику для представления движения тел. В этом разделе, употребляя слово «движение», я подразумеваю поворот или перенос твердого тела; исключается любого вида деформация тела или его части, будь то растяжение, изгиб или сжатие. В следующей главе будут рассматриваться и такие деформации.

Заглавными буквами мы будем обозначать точки, принадлежащие телам, как это принято в геометрии. Маленькие буквы будут обозначать различного рода движения. Например, мы можем сказать: пусть x означает «передвинут на один сантиметр на восток», а t — «повернут на 90° вокруг фиксированной точки O». Равенство

[1] От англ. boiled — вареный.— *Прим. перев.*
[2] От англ. red — красный.— *Прим. перев.*
[3] От англ. dead — мертвый.— *Прим. перев.*

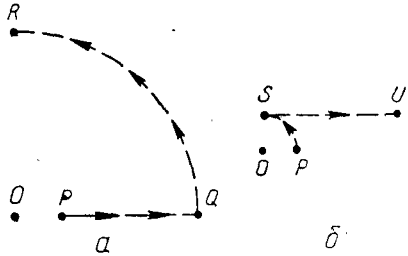


Рис. 31.

вида $Q=xP$ будет означать, что Q есть точка, в которую вы попадаете, переместив P на один сантиметр к востоку.

Мы можем рассмотреть и результат двух последовательных движений. Предположим, мы берем точку P, перемещаем ее на один сантиметр на восток и получаем Q, а затем поворачиваем Q на 90° вокруг точки O, тогда наша точка окажется в положении R (рис. 31*а*). Символически мы можем записать это гораздо короче, а именно: $Q=xP$, $R=tQ$. Поскольку $Q=xP$, то вполне естественно заменить во втором равенстве Q на xP и написать $R=txP$. Это показывает, что R получается из P путем применения сначала операции x, а потом операции t. Суммарное действие обозначается tx. Заметьте, что операция, выполняемая второй, стоит на первом месте. Это хорошо согласуется с порядком слов в обычной разговорной речи. «Он покинул горящий дом» означает, что дом сначала загорелся, а потом он его покинул. (Некоторые авторы, занимающиеся теорией групп, принимают другой порядок букв, и этот пустяковый вопрос является причиной большой путаницы.)

Порядок выполнения операций не безразличен. На рис. 31*б*, показано, что означает xt: $S=tP$, $U=xS$; поэтому $U=xtP$. Точка U не то же самое, что точка R; поэтому воздействие xt отличается от воздействия tx, т. е. $xt \neq tx$ (читается как «xt не равно tx»).

Рассматривая алгебру логики и алгебру классов, мы заметили, что там ab имеет тот же смысл, что ba. Когда $ab=ba$, говорят, что умножение *коммутативно*. Но если порядок букв играет роль, то мы говорим, что умножение *некоммутативно*. Очень часто умножение бывает некоммутативно, поэтому необходимо привыкнуть различать порядок букв. Хорошо известный пример мы находим в дифференциальном исчислении, где $x\frac{d}{dx}y$ не то же самое, что $\frac{d}{dx}xy$. Другой замечательный пример мы находим в квантовой теории: $pq-qp=\frac{h}{2\pi i}$. Ничего загадочного в этом уравнении нет; p и q — это не числа, а обозначения для операций, и оказывается, что порядок их выполнения играет существенную роль — только и всего.

ГРУППА ПРЯМОУГОЛЬНИКА

Предположим, что у нас есть прямоугольник, вырезанный из картона, и мы чертим на листе бумаги рамку, вмещающую этот прямоугольник (рис. 32). Это легче всего сделать, положив картонный

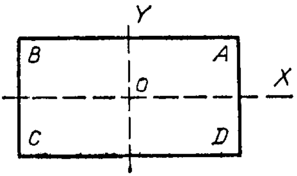


Рис. 32.

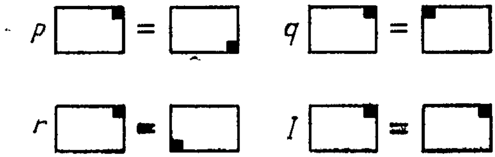


Рис. 33.

прямоугольник на бумагу и обведя его карандашом. Я могу совершить с прямоугольником несколько операций, которые оставляют его лежащим в рамке:

1) я могу повернуть его вокруг оси OX, так что поменяются местами точки A и D, а также точки B и C — назовем эту операцию p;

2) могу повернуть его вокруг оси OY — операция q;

3) могу повернуть картон на 180° вокруг точки O, не отрывая его от листа бумаги, тогда местами поменялись бы точки A и C, а также B и D — операция r;

4) могу просто оставить прямоугольник в том положении, в каком он есть,— операция I.

Если мы для наглядности зачерним один из углов прямоугольника (картон надо закрасить с обеих сторон), мы получим то, что изображено на рис. 33, показывающем результат воздействия операций p, q, r, I [1].

Итак, поместить прямоугольник в рамку можно только четырьмя способами. Черный угол должен попасть в один из четырех углов, и после того, как вы решили, в какой именно угол, имеется лишь один способ сделать это.

Если мы выполняем последовательно две операции, то наш прямоугольник окажется все равно в одном из четырех указанных выше положений. Например, мы хотим знать, к чему приведет операция qp. В исходном положении прямоугольника пятно находится в северо-восточном углу. Операция p отправляет его на юго-восток. Затем q посылает наше пятно в юго-западный угол. Но к такому результату могла бы привести единственная операция r. Значит, $qp=r$. Точно так же вы можете проверить, что $pp=I$; этот результат обычно записывается как $p^2=I$. Столь же легко вы получите и остальные зависимости. Все они собраны в следующей таблице умножения.

---

[1] Символ I используется здесь отчасти потому, что он является начальной буквой слова identical (идентичный), отчасти оттого, что он напоминает цифру 1, с которой у него много схожих свойств. Чтобы быть последовательным, нужно было использовать для этой операции строчную букву. Но i у нас всегда ассоциируется с $\sqrt{-1}$, так что лучше уж быть непоследовательным.

Операции, выполняемые в первую очередь (записываются на втором месте)

Операции, выполняемые во вторую очередь

	I	p	q	r
I	I	p	q	r
p	p	I	r	q
q	q	r	I	p
r	r	q	p	I

Составляя таблицу умножения, обычно необходимо указать, какие операции выполняются первыми, так как, вообще говоря, ab и ba — разные вещи. Это сделано в нашей таблице, хотя в конце концов в данном частном случае это оказывается ненужным; например, и pq и qp равны r. Вообще, как видно из таблицы, произведение любых двух из этих трех букв равно третьей. Такая таблица умножения коммутативна.

Но вообще следует ожидать, что умножение некоммутативно, как, скажем, в следующем примере.

ГРУППА РАВНОСТОРОННЕГО ТРЕУГОЛЬНИКА

Точно так же, как мы сделали рамку для прямоугольника, мы можем взять картонный равносторонний треугольник, обвести его карандашом на бумаге и рассмотреть все способы его размещения в рамке (рис. 34).

Имеется шесть возможных движений: 1) можно оставить треугольник как он есть — операция I; 2) можно повернуть его на 120° против часовой стрелки — операция ω; 3) можно повернуть треугольник на 240°; это все равно, что дважды выполнить операцию ω, поэтому назовем такой поворот ω^2; 4) можно повернуть треугольник вокруг пунктирной линии *1* — операция p; 5) можно повернуть его вокруг пунктирной линии *2* — операция q; 6) можно повернуть треугольник вокруг линии *3*—операция r.

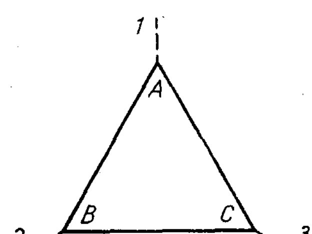


Рис. 34.

Если треугольник находится в положении $\overset{A}{B\quad C}$, то операция I оставляет его в том же положении. Операция ω перемещает его в положение $\overset{C}{A\quad B}$. Если бы мы использовали операцию ω^2, то треугольник оказался бы в положении $\overset{B}{C\quad A}$. Подобным же образом операция p переводит его из положения $\overset{A}{B\;C}$ в $\overset{A}{C\;B}$, операция q — из $\overset{A}{B\;C}$ в $\overset{C}{B\;A}$, операция r — из $\overset{A}{B\;C}$ в $\overset{B}{A\;C}$.

Имеется только шесть способов, которыми можно разместить три буквы в углах треугольника, так что перечисленные выше положения должны включать все возможные способы размещения треугольника в рамке. Если мы выполним последовательно две операции, например. сначала r, а затем p, то мы неизбежно приведем треугольник к одному из указанных выше шести положений; тем самым последовательное выполнение двух операций дает тот же результат, что и выполнение одной-единственной операции из шести, перечисленных выше. Так, например, pr оказывает на $\overset{A}{B\;C}$ такое же действие, как ω^2, ибо в обоих случаях мы получаем $\overset{B}{C\;A}$. Поэтому $pr = \omega^2$. Если мы выполняем операции в обратном порядке, т. е. если мы рассматриваем rp, то мы приходим к другому результату: rp меняет положение $\overset{C}{A\;B}$ на $\overset{A}{B\;C}$, и то же самое делает операция ω. Поэтому $rp = \omega$. Возможно, вам самим захочется дополнить таблицу умножения, прежде чем проверять таблицу, приводимую ниже.

Операции, выполняемые в первую очередь (записываются на втором месте)

Операции, выполняемые во вторую очередь

	I	ω	ω^2	p	q	r
I	I	ω	ω^2	p	q	r
ω	ω	ω^2	I	r	p	q
ω^2	ω^2	I	ω	q	r	p
p	p	q	r	I	ω	ω
q	q	r	p	ω^2	I	ω^2
r	r	p	q	ω	ω^2	I

ДРУГИЕ ГРУППЫ ДВИЖЕНИЙ

Рассмотрев другие фигуры, можно найти еще ряд групп движений. Чем симметричнее фигура, тем больше количество возможных движений в группе. В качестве примера можно рассмотреть очертания букв алфавита. Вырезанная из картона буква S может быть уложена в свой контур только двумя способами. Ее группа состоит из операции I и из поворота на 180° относительно центра. Если мы назовем последнюю операцию k, то таблица умножения будет иметь вид:

	I	k
I	I	k
k	k	I

Некоторые буквы вообще не обладают свойством симметрии, это буквы вроде F, G, K, Q. Такие буквы, помещенные в свою рамку, никак не могут быть перемещены в ней, не считая, конечно, того, что их можно просто вынуть и положить снова. Группа состоит из единственной операции I, и таблица умножения тривиальна:

	I
I	I

Анализируя эти таблицы умножения, можно заметить очень много интересных вещей. Группы — очень важный объект в математике. Они входят в удивительно широкий круг математических исследований. С помощью теории групп можно получить самые неожиданные, не связанные друг с другом результаты. Так, можно показать, что уравнение пятой степени неразрешимо алгебраическими методами, что существует всего лишь 17 существенно различных типов обоев, а также получить важные сведения о структуре молекул в химии. Одно время математики даже считали, что группы являются ключом к раскрытию всех тайн вселенной, и за это трудно их винить.

В этой главе я хотел обратить ваше внимание на само существование теории групп и подчеркнуть, что математические символы не обязательно представляют числа; они могут обозначать и такие слова, как «или», или операции, вроде «повернуть на 120°». В восьмой главе мы будем рассматривать матрицы, которые можно считать особого рода операциями. Они включают как рассмотренные нами выше жесткие движения, так и другие типы перемещений, при которых допускается деформация фигуры.

Глава восьмая

МАТРИЧНАЯ АЛГЕБРА

Многое можно сказать об этой теории матриц, которая, как мне кажется, должна предшествовать теории определителей.

А. Кэли, 1855

Декарт, о котором мы уже упоминали, показал, что любой геометрический результат может быть представлен в алгебраическом виде. Точки геометрических фигур можно представить себе нанесенными на координатную сетку; тогда положение каждой точки определится парой чисел (x, y). Любое свойство, встречающееся в геометрической теореме, может быть переведено на язык алгебраических соотношений между координатами x и y различных точек.

Некоторые математики в восторге от этого изобретения Декарта, поскольку оно позволяет упразднить геометрию как предмет и заменить его алгеброй. Другие же предпочитают мыслить геометрическими образами без обращения к алгебре. Но, как мне кажется, истинная ценность идеи Декарта заключается в том, что она позволяет постоянно переходить от алгебры к геометрии и обратно. Очень часто смысл алгебраического результата виден гораздо лучше, если перевести его на язык геометрии, поскольку геометрия позволяет увидеть и почувствовать алгебраические абстракции. С другой стороны, геометрические результаты становятся более точными и ясными, когда они принимают алгебраическую или арифметическую форму.

Глава седьмая называлась «Алгебра без арифметики». В последнем разделе этой главы мы рассматривали такие геометрические понятия, как «поворот на 120°» или «перемещение на 3 сантиметра на восток». Даже в этих геометрических понятиях встречаются цифры, вроде только что упомянутых 120 или 3. Сейчас мы собираемся перевести эти геометрические понятия почти полностью на язык чисел. Нам не нужно будет тогда заявлять, что такая-то операция есть поворот; цифры будут говорить сами за себя, и из них нам станет ясно, что данная операция есть действительно поворот. Для того чтобы определить операцию, нам понадобится четыре числа. Так опять появляется арифметика. Но нужно отчетливо понять, что предметом обсуждения будут все-таки операции, а не числа. Например, силу удара можно описать, сказав, что он ощущается, как если бы вас толкнули две тонны свинца, несущиеся со скоростью 100 *км* в час. Мы могли бы выразить это символически, сказав, что это был удар (2, 100). Но мы имеем здесь не просто пару чисел; запись (2, 100) характеризует удар, и обсуждать нужно именно удар, а не просто эту запись. Итак, в этой главе числа будут определять операции, действия, движения; поэтому, производя вычисления, всегда следует помнить об операциях, которые они собой представляют. Имея это в виду, легко составить правила, по которым следует проводить вычисления. После того как эти правила сформулированы, можно представить предмет абстрактно, хотя бы ради ангела, с которым мы связаны по телефону. Таким образом, предмет принял бы чисто арифметическую форму. Это, конечно, было бы очень удобно с точки зрения логической точности, но серьезно помешало бы пониманию и оценке предмета в целом.

ОПЕРАТОРЫ ОТРАЖЕНИЯ

Операции, рассмотренные в конце седьмой главы, были двух типов: либо фигура поворачивалась вокруг некоторой прямой, либо вращалась относительно точки. Сейчас мы собираемся рассмотреть эти операции в духе Декарта.

В седьмой главе мы обозначили операции малыми буквами, с тем чтобы отличать их от геометрических точек P, Q и т. п. В этой главе малые буквы (x, y и т. п.) будут обозначать числа; точки мы по-прежнему обозначаем большими буквами. Что же касается операторов, то они, во избежание путаницы, будут обозначаться заглавными буквами, выделенными жирным шрифтом: **A**, **B**, **C** и т. д.

Рассмотрим сначала операцию **A**, которая означает «отражение (поворот) относительно оси OX».

На следующих далее рисунках P_0 означает положение точки до того, как была проделана соответствующая операция, а P_1 — положение точки после операции; (x_0, y_0) — координаты точки P_0, (x_1, y_1) — координаты точки P_1.

Для операции **A**, как это явствует из рис. 35, точка P_1 имеет ту же абсциссу, что и точка P_0, а ординаты этих точек отличаются только знаками. Поэтому мы имеем:

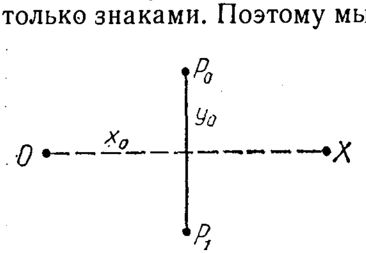


Рис. 35.

$$\mathbf{A}\begin{cases} x_1 = x_0 \\ y_1 = -y_0 \end{cases}$$

Эти два уравнения полностью определяют операцию **A**. Они показывают, что происходит с любой точкой, имеющей координаты (x_0, y_0).

Подобным же образом, если операция **B** означает «отражение относительно оси OY», мы имеем (рис. 36):

$$\mathbf{B}\begin{cases} x_1 = -x_0 \\ y_1 = y_0 \end{cases}$$

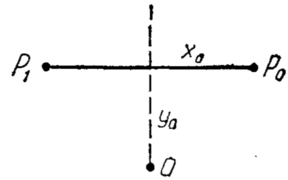


Рис. 36.

Обозначая буквой **C** «поворот на 180° вокруг точки O», мы находим (рис. 37):

$$\mathbf{C}\begin{cases} x_1 = -x_0 \\ y_1 = -y_0 \end{cases}$$

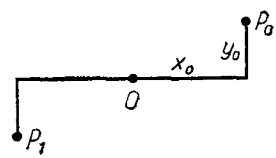


Рис. 37.

Операции **A**, **B**, **C** встречались нам в главе седьмой в группе прямоугольника; там они назывались *p, q, r*.

Мы можем рассмотреть еще операцию **D**, где вместо отражения относительно оси OX или OY производится отражение относительно прямой OL, составляющей с осью OX угол α. Это отражение переводит P_0 в P_1. Если N_0 есть точка на оси OX, лежащей непосредственно под P_0, то отражением точки N_0 является точка N_1. Мы знаем, что длина ON_1 равна ON_0, т. е. x_0, а длина N_1P_1 равна N_0P_0, т. е. y_0. Известно нам и направление ON_1, оно составляет угол 2α с OX; N_1P_1 — перпендикуляр к ON_1. Следовательно, мы можем найти координаты P_1, отправившись от точки O к P_1 через N_1. Результат выглядит следующим образом (рис. 38):

$$\mathbf{D}\begin{cases} x_1 = x_0\cos 2\alpha + y_0 \sin 2\alpha \\ y_1 = x_0 \sin 2\alpha - y_0\cos 2\alpha \end{cases}$$

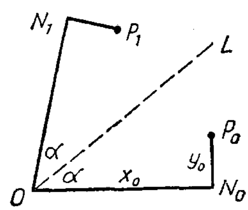


Рис. 38.

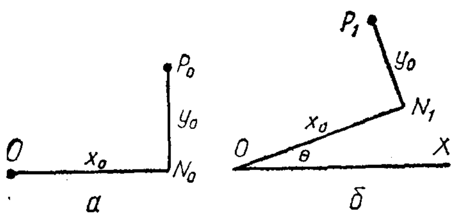
В частном случае, когда $\alpha=45°$, операция упрощается. Назовем ее **E**:

$$\mathbf{E}\begin{cases} x_1 = y_0 \\ y_1 = x_0 \end{cases}$$

E означает отражение относительно прямой $y=x$. Эффект операции **E** заключается в простой перестановке координат точки. Так, если P_0 — точка с координатами (3, 4), то $P_1=\mathbf{E}P_0$ будет иметь координаты (4, 3).

ПОВОРОТЫ

Очень сходные вычисления дают нам возможность определить операцию поворота. Пусть **F** — операция поворота на угол θ относительно центра O. На рис. 39*а* показано положение точки P_0. Поворачиваем прямые линии этого рисунка на угол θ, как показано на рис. 39*б*, и получаем точку P_1. Проследив путь от точки O к точке P_1 через N_1, находим:

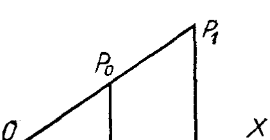


$$\mathbf{F}\begin{cases} x_1 = x_0\cos\theta - y_0\sin\theta \\ y_1 = x_0\sin\theta + y_0\cos\theta \end{cases}$$

Рис. 39.

РАСТЯЖЕНИЯ

Прежде чем сделать общий вывод из всех приведенных формул, рассмотрим еще одну операцию. Эта операция называется растяжением (рис. 40); обозначим ее буквой **G**. Здесь точка P_0 переходит в точку P_1, лежащую на прямой OP_0, причем расстояние OP_1 в k раз превосходит OP_0, где k — некоторое постоянное число. Операция заключается просто в изменении масштаба чертежа. Из сравнения подобных треугольников усматриваем:



$$\mathbf{G}\begin{cases} x_1 = kx_0 \\ y_1 = ky_0 \end{cases}$$

Рис. 40.

МАТРИЧНАЯ СИМВОЛИКА

Рассматривая все полученные результаты, мы видим, что в каждом случае x_1 и y_1 представлены выражениями, линейно зависящими от x_0 и y_0. Это значит, что в каждом случае мы имеем уравнения вида

$$x_1 = ax_0 + by_0,$$
$$y_1 = cx_0 + dy_0.$$

Можно определить таким образом любую из рассмотренных операций, указав, чему равны для нее числа a, b, c и d. Для определения операций принято записывать числа a, b, c и d в том порядке, в котором они встречаются в уравнениях, а именно:

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

Этот способ известен как матричный метод определения операций. Группа чисел, расположенных таким образом, называется *матрицей*[1].

Каждая из рассмотренных ранее операций может быть записана в виде матрицы. Правда, в ряде формул отсутствуют некоторые члены; так, например, в выражении для операции **E** в правой части стоят лишь два числа, вместо возможных четырех. Будем считать, что отсутствующие члены имеют нулевые коэффициенты. Тогда формула для операции **E** может быть записана как

$$\mathbf{E} \begin{cases} x_1 = 0x_0 + y_0 \\ y_1 = x_0 + 0y_0 \end{cases}$$

В соответствии с этим в матричном обозначении:

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

Точно так же мы можем записать и другие операции:

$$\mathbf{A} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \qquad \mathbf{B} = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}, \qquad \mathbf{C} = \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix},$$

$$\mathbf{D} = \begin{pmatrix} \cos 2\alpha & \sin 2\alpha \\ \sin 2\alpha & -\cos 2\alpha \end{pmatrix}, \quad \mathbf{F} = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}, \quad \mathbf{G} = \begin{pmatrix} k & 0 \\ 0 & k \end{pmatrix}.$$

В некоторых случаях бывает полезно и символы x_1, y_1, x_0, y_0 записывать в виде матрицы, наподобие a, b, c, d. Чтобы показать, что

---

[1] В полиграфии матрицей называется форма, в которую печатник отливает шрифт. В математике же матрица — это просто таблица, в которой размещены числа.

некоторая матрица действует на точку с координатами x_0, y_0, мы пишем после этой матрицы

$$\begin{pmatrix} x_0 \\ y_0 \end{pmatrix}$$

Итак, в окончательном виде симзолическое выражение для операции **E** запишется следующим образом:

$$\begin{pmatrix} x_1 \\ y_1 \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} x_0 \\ y_0 \end{pmatrix}$$

Нельзя сказать, чтобы такая запись была особенно экономной; однако, как мы скоро выясним, она имеет свои преимущества. Конечно, самой компактной формой записи этого утверждения явилась бы его первоначальная геометрическая форма, а именно: $P_1 = \mathbf{E}P_0$.

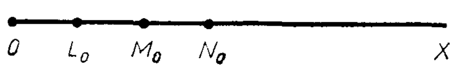
МАТРИЦА В ОБЩЕМ ВИДЕ

Самый общий вид матрицы появляется в уравнениях

$$\begin{aligned} x_1 &= ax_0 + by_0 \\ y_1 &= cx_0 + dy_0 \end{aligned} \quad \text{или} \quad \begin{pmatrix} x_1 \\ y_1 \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} x_0 \\ y_0 \end{pmatrix}.$$

Мы знаем, что при определенных значениях коэффициентов a, b, c и d эти уравнения представляют операции поворота, отражения или растяжения. Но только ли эти операции могут быть представлены с помощью данного общего выражения? Как воздействует на геометрическую фигуру произвольная матрица? Как это действие изобразить наглядно?

Исследуем этот вопрос для совершенно произвольных значений коэффициентов a, b, c и d. Мы начнем с того, что посмотрим, что происходит с точками, расположенными на оси OX.



Возьмем точки L_0, M_0, N_0 на расстояниях 1, 2, 3 от точки O и подставим их координаты в приведенные выше уравнения. Поскольку L_0 — точка с координатами (1, 0), то L_1 должна иметь координаты (a, c). Аналогично для точек M_0 и N_0: так как координаты точки M_0 равны (2, 0), то M_1 должны иметь координаты $(2a, 2c)$; так как точка N_0 имеет координаты (3, 0), то координаты точки N_1 будут $(3a, 3c)$.

Нанеся точки на координатную сетку (рис. 41), мы видим, что они лежат на прямой линии на равных расстояниях друг от друга.

Точно так же, располагая точки P_0, Q_0, R_0 на оси OY, мы находим, что (0, 1) переходит в (b, d), (0, 2) — в $(2b, 2d)$, (0, 3) — в $(3b, 3d)$ (рис. 42).

Какое же положение займет после операции точка S_0 с координатами (1, 1)? Полагая $x_0=1$ и $y_0=1$, мы находим, что S_1 имеет координаты $(a+b,\ c+d)$ (рис. 43).

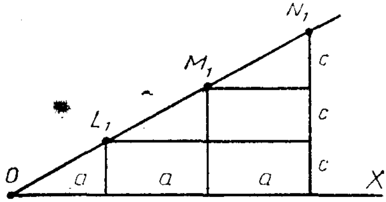


Рис. 41.

Нетрудно показать, что четырехугольник $OL_1S_1P_1$ является параллелограммом. Вершины этой фигуры имеют координаты (0, 0), $(a,\ c)$, $(a+b,\ c+d)$, $(b,\ d)$. Легко вычислить наклоны сторон и убедиться, что противоположные стороны имеют одинаковый наклон [1].

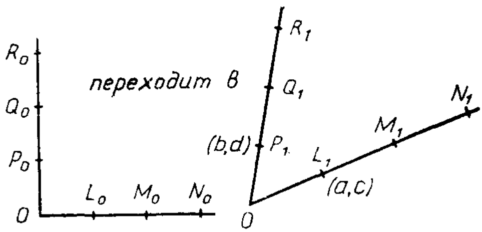


Рис. 42.

И действительно, как показывает дальнейшее небольшое исследование, прямые линии переходят в прямые, параллелограммы — в параллелограммы, начало координат остается на месте (рис. 44).

Следовательно, результат действия матрицы $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ можно видеть без каких-либо вычислений. Два числа в первом столбце a и c дают координаты точки L_1. Числа b и d, стоящие во втором столбце, являются координатами точки P_1. На сторонах OL_1 и OP_1 можно построить параллелограмм и, повторяя это построение, получить всю картину.

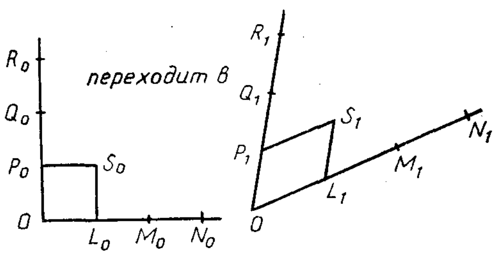


Рис. 43.

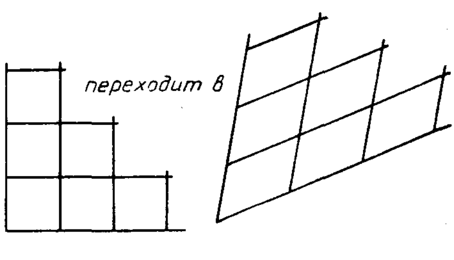


Рис. 44.

[1] Этот результат, конечно, хорошо известен студентам, изучающим теоретическую механику. Он выражает тот факт, что равнодействующая, найденная при помощи построения параллелограмма сил, равна равнодействующей, найденной сложением вертикальной и горизонтальных составляющих.

И обратно, если такая решетка уже начерчена, мы можем всегда найти матрицу, для которой эта решетка построена. Мы просто считываем координаты точек L_1 и P_1 и записываем их в первом и во втором столбцах матрицы.

Например, пусть нужно составить матрицу для операции поворота на 90°. При таком повороте точка L_0 переходит в положение (0, 1), а P_0 — в положение (—1,0). Нам потребуются столбцы

$$\begin{matrix} 0 \\ 1 \end{matrix} \quad \text{и} \quad \begin{matrix} -1 \\ 0 \end{matrix},$$

а вся матрица будет иметь вид:

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

ПРИМЕНЕНИЯ МАТРИЦ

Путь, которым мы подошли к идее матрицы, может навести на мысль, что матрицы полезны только при изучении поворотов, отражений и других преобразований геометрических фигур. Однако это не так. Матрицы проникли во все области математики.

Некоторые применения связаны непосредственно с той областью, о которой мы уже говорили, — с геометрическими деформациями. Наглядным примером применения матриц является строительная механика. Если некоторый материал, будь то резина или сталь, подвергается нагрузке, всегда происходит изменение формы, *деформация*. Деформация может быть очень незначительной, но она всегда есть. Даже стальной брусок не может оказывать давление или передавать его, пока он сохраняет свою первоначальную форму. (Это свойство связано с законом Гука. Мы можем рассматривать молекулы стали как совокупность крошечных шариков, связанных друг с другом при помощи чрезвычайно сильных пружин.) Воздействие деформации приводит к тому, что маленькие квадраты материала превращаются (с достаточной степенью приближения) в маленькие параллелограммы; поэтому деформация может быть описана с помощью матрицы.

В теории электричества и магнетизма также встречаются матрицы, поскольку тело, находящееся под воздействием электрических и магнитных сил, ведет себя в известном смысле как материал при деформации.

Матрицы встречаются и в аэродинамике, при изучении воздушного потока, обтекающего крыло самолета. Представьте себе квадратик из дыма, затянутый в воздушный поток; форма квадратика будет изменяться по мере его продвижения. Способ изменения формы покажет, что происходит в потоке. Естественно, что для описания этого процесса используются матрицы. В книгах по аэродинамике вы встречаетесь с понятием «безвихревой поток», которое означает, что квадратики деформируются, но не вращаются. (Здесь нужна некоторая осторожность, поскольку это понятие нуждается в точном определении.) Тот же самый термин — «безвихревой» — используется и в гидродинамике (науке о движении жидкости).

Во всех этих примерах имеется прямая связь с геометрией. Но существует очень много случаев применения матриц, где геометрия не является связующим звеном. Связь достигается благодаря *алгебраической форме* матриц. Матрицы, рассмотренные ранее в этой главе, являлись способом записи *линейных уравнений*. Но линейные уравнения, естественно, могут встречаться в любой области математики. (Существует мнение, в значительной степени верное, что математика достигла больших успехов в решении только тех задач, которые в каком-то смысле линейны. Нелинейные задачи содержат большие трудности, которые до сих пор не умеют разрешать.) Там, где встречаются линейные уравнения, всегда потребуются матрицы.

Линейные уравнения автоматически возникают в задачах, касающихся *малых величин*. Предположим, что у нас есть функция $f(x, y)$, и нас интересует, как она ведет себя в небольшой области около начала координат, т. е. вблизи точки $x=0$, $y=0$. Если $f(x, y)$ принадлежит к некоторому широкому классу функций, то оказывается возможным разложить ее в ряд по степеням x и y, скажем,

$$f(x, y) = k + ax + by + ex^2 + gxy + hy^2 + mx^3 + nx^2y + \dots .$$

Но x и y малы. Поэтому x^2, xy, y^2 и т. д. и подавно малы, т. е. ими можно пренебречь. С этой степенью приближения

$$f(x, y) = k + ax + by.$$

При $x = y = 0$ имеем $f(x\ y) = k$. Значит, $k = f(0, 0)$. Отсюда следует, что

$$f(x, y) - f(0,0) = ax + by.$$

Выражение, стоящее в левой части, представляет приращение функции при переходе от точки $(0, 0)$ к точке (x, y). Справа же стоит $ax+by$, т. е. как раз выражение того типа, которое привело нас раньше к матрицам. Но чтобы составить матрицу, нам необходимо иметь два таких выражения. Поэтому в задачах, касающихся изменения двух функций $f(x, y)$ и $g(x, y)$, — а таких задач много, — всегда встретятся матрицы.

Важным инженерным применением матриц является теория малых колебаний тел. Матрицы применяются там потому, что, как уже объяснялось выше, колебания *малы*. В любой машине могут произойти серьезные повреждения, если частота работы двигателя совпадет с собственной частотой колебания какой-нибудь другой детали машины. При определенных условиях постоянные вибрации могут даже привести к серьезным поломкам. Хорошо известно, что солдаты, проходя по мосту, сбивают шаг. (Будучи студентами последнего курса в Кембридже, некоторые из нас частенько простукивали фонарные столбы в надежде найти их собственную частоту колебаний и заставить их развалиться. Правда, нам ни разу это не удалось.)

Развитие квантовой теории многим обязано матричному исчислению (Гейзенберг, Дирак). Широко используемые в теории относительности тензоры являются обобщением матриц. Матрицы встречаются также во многих разделах чистой математики. Простейшим примером являются конические сечения; проективная геометрия, теория групп, дифференциальные уравнения — это уже следующие, более серьезные этапы.

Матричная алгебра — наиболее убедительный пример того, как одна и та же закономерность встречается при самых различных обстоятельствах.

УМНОЖЕНИЕ МАТРИЦ

Поскольку буквами **A** и **B** мы обозначаем геометрические операции, то под **AB** мы будем подразумевать матрицу, которая представляет результат последовательного выполнения операций **B** и **A**.

Если мы рассмотрим произвольную точку P_0 [на большей части предыдущих рисунков P_0 лежала на оси OY и имела координаты $(0,1)$], то в результате операции **B** эта точка перейдет в положение P_1:

$$P_1 = \mathbf{B}P_0.$$

Если далее выполняется операция **A**, то точка P_1 перейдет в положение P_2:

$$P_2 = \mathbf{A}P_1.$$

Комбинируя оба результата, мы записываем аналогично тому, как мы делали это в седьмой главе:

$$P_2 = \mathbf{AB}P_0.$$

Везде в этой главе мы будем употреблять цифры 0, 1, 2 именно в этом смысле. Индекс 0 будет употребляться для обозначения первоначального положения точки, 1 указывает промежуточное положение точки после выполнения первой операции, 2 относится к положению точки после двух операций.

Беря в качестве **A** и **B** те же самые операции, что и в начале этой главы, мы можем записать из рассмотрения группы прямоугольника, что $\mathbf{AB} = \mathbf{C}$.

Этот результат можно получить и алгебраическим путем. $P_1 = \mathbf{B}P_0$ означает, что

$$\begin{aligned} x_1 &= -x_0 \\ y_1 &= y_0, \end{aligned}$$

а $P_2 = \mathbf{A}P_1$ — что
$$x_2 = x_1$$
$$y_2 = \quad -y_1.$$

Из этих уравнений и ребенку будет нетрудно вывести следующие:
$$x_2 = -y_0$$
$$y_2 = \quad -y_0.$$

В матричных обозначениях это можно записать следующим образом:

из
$$\begin{pmatrix} x_2 \\ y_2 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ y_1 \end{pmatrix}$$

и
$$\begin{pmatrix} x_1 \\ y_1 \end{pmatrix} = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} x_0 \\ y_0 \end{pmatrix}$$

следует
$$\begin{pmatrix} x_2 \\ y_2 \end{pmatrix} = \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} x_0 \\ y_0 \end{pmatrix}.$$

Но мы можем также объединить уравнения «из... и...» и тогда получим:
$$\begin{pmatrix} x_2 \\ y_2 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} x_0 \\ y_0 \end{pmatrix}.$$

Итак, мы видим, что
$$\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix},$$
т. е. $\mathbf{AB} = \mathbf{C}$.

Метод, которым мы действовали в данном случае, носит общий характер. Если у нас есть любые две матрицы $\mathbf{U}$ и $\mathbf{V}$, мы можем найти $\mathbf{UV}$ следующим образом. Мы записываем уравнения $P_2 = \mathbf{U}P_1$, $P_1 = \mathbf{V}P_0$ в их алгебраической форме; другими словами, мы пишем уравнения, выражающие зависимость x_2 и y_2 от x_1 и y_1, а также x_1 и y_1 от x_0 и y_0. Далее мы подставляем в первую пару уравнений вместо x_1 и y_1 их выражения через x_0 и y_0 (взятые из второй пары уравнений); в результате мы получаем два уравнения, с помощью которых x_2, y_2 выражаются прямо через x_0, y_0. Эти новые уравнения представляют соотношение $P_2 = \mathbf{W}P_0$; матрица $\mathbf{W}$ составлена определенным образом из коэффициентов уравнений для $\mathbf{U}$ и $\mathbf{V}$:
$$\mathbf{W} = \mathbf{UV}.$$

Я собираюсь подробно все это проделать, с тем чтобы выработать правило умножения матриц. Пусть
$$\mathbf{U} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}, \text{ а } \mathbf{V} = \begin{pmatrix} p & q \\ r & s \end{pmatrix},$$

где a, b, c, d, p, q, r, s — конечно, просто числа. Уравнение $P_2 = \mathbf{U}P_1$ в развернутом виде представляется парой уравнений

$$x_2 = ax_1 + by_1,$$
$$y_2 = cx_1 + dy_1,$$

а уравнение $P_1 = \mathbf{V}P_0$ —

$$x_1 = px_0 + qy_0,$$
$$y_1 = rx_0 + sy_0.$$

Подставляя в первую пару уравнений значения x_1 и y_1, выраженные через x_0 и y_0, получим:

$$x_2 = a(px_0 + qy_0) + b(rx_0 + sy_0) = (ap + br)x_0 + (aq + bs)y_0,$$
$$y_2 = c(px_0 + qy_0) + d(rx_0 + sy_0) = (cp + dr)x_0 + (cq + ds)y_0.$$

Величины, являющиеся коэффициентами при x_0 и y_0 в последней паре уравнений, дают числа, которые должны быть вставлены в матрицу $\mathbf{W}$. Таким образом,

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}\begin{pmatrix} p & q \\ r & s \end{pmatrix} = \begin{pmatrix} ap + br & aq + bs \\ cp + dr & cq + ds \end{pmatrix}.$$

Это можно рассматривать как формулу для умножения двух матриц. Вы можете найти по этой формуле произведение любых двух матриц, подставляя значения чисел a, b, c, d, p, q, r, s. Но на практике существует лучшее правило. Если вы взглянете на член $ap + br$ в произведении, то увидите, что буквы a и b взяты из верхней строки первой матрицы, а буквы p и r — из левого столбца второй матрицы. Отбросив все остальное, мы увидим только

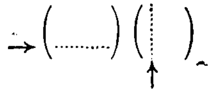
$$\begin{array}{cccc} \to\begin{pmatrix} a & b \\ \cdot & \cdot \end{pmatrix} & \begin{pmatrix} p & \cdot \\ r & \cdot \end{pmatrix} & = & \to\begin{pmatrix} ap + br & \cdot \\ \cdot & \cdot \end{pmatrix}. \\ & \uparrow\quad & & \uparrow\qquad \end{array}$$

Аналогичным образом получаются и другие члены произведения. Если мы нарисуем стрелку против любой строки первой матрицы и другую стрелку против какого-нибудь из столбцов второй матрицы, то числа, выделенные таким образом, составят одно выражение в матрице произведения. Место, где стоит это выражение, показано двумя стрелками, нарисованными рядом с матрицей произведения в положениях, соответствующих стрелкам в левой части равенства.

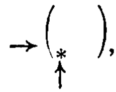
Так, числа, стоящие в верхней строке и правом столбце, появятся в матрице произведения в том месте, где верхняя строка пересекается с правым столбцом:

$$\begin{array}{cccc} \to\begin{pmatrix} \cdots\cdots \\ \end{pmatrix} & \begin{pmatrix} & \vdots \\ & \vdots \end{pmatrix} & = & \to\begin{pmatrix} & * \\ & \end{pmatrix} \\ & \quad\uparrow & & \quad\uparrow \end{array}$$

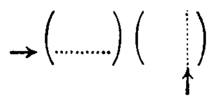
Подобным же образом



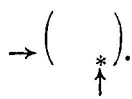
дает член произведения



а



дает



Все это гораздо легче объяснить на доске, чем описать в книге.

Я думаю, что вы все-таки усмотрите правило для составления членов произведения. Возьмем наш первый пример. В верхней строке стояли числа a, b, а в левом столбце — p, r. Умножаем соответствующие числа и получаем ap, br; затем складываем эти произведения и окончательно имеем: $ap+br$.

Всякий, кому приходится много работать с матрицами, настолько осваивается с этим правилом умножения, что он мог бы умножать во сне. Палец левой руки автоматически скользит вдоль строк первой матрицы, палец правой руки опускается вдоль столбцов второй матрицы. В этот момент человек производит умножение, а затем записывает результат. Остается только пожалеть, что природа не снабдила нас третьей рукой.

Если вы проделаете несколько примеров, то убедитесь, что к этому правилу очень легко привыкнуть. Лучшим упражнением служат те примеры, для которых вы заранее знаете ответ, тогда вы сразу обнаружите любую ошибку.

Например, рассмотренные нами в этой главе матрицы **A**, **B**, **D**, **E** представляют операции отражения. Операция отражения, произведенная дважды, возвращает нас в исходное положение. Пользуясь обозначением $\mathbf{A}^2$ для $\mathbf{AA}$ мы могли бы найти, что $\mathbf{A}^2$, $\mathbf{B}^2$, $\mathbf{D}^2$ и $\mathbf{E}^2$ равны **I**, где I означает то же самое, что и в главе седьмой, т. е. операцию, оставляющую все на своих местах. Найдите сами матрицу **I** — для операции I.

Далее, поворот на угол α, а затем на угол β — это все равно, что поворот на угол $\alpha+\beta$. Поэтому, используя матрицу **E** для операции поворота, мы могли бы найти, что

$$\begin{pmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{pmatrix} \begin{pmatrix} \cos\beta & -\sin\beta \\ \sin\beta & \cos\beta \end{pmatrix} = \begin{pmatrix} \cos(\alpha+\beta) & -\sin(\alpha+\beta) \\ \sin(\alpha+\beta) & \cos(\alpha+\beta) \end{pmatrix}.$$

Этот результат может служить одним из способов доказательства формул для косинуса и синуса суммы двух углов.

Возможно, вы захотите сравнить **FD** и **DF**. Тогда вы увидите, что это различные операции. Оба произведения являются операциями отражения; первое — относительно прямой, составляющей с осью x угол $\alpha + \frac{1}{2}\theta$, а второе — относительно прямой, составляющей с осью x угол $\alpha - \frac{1}{2}\theta$. Каждое произведение должно поэтому иметь вид матрицы **D**, где угол α заменен одним из только что упомянутых углов. Справедливость этих утверждений можно проверить геометрически.

СЛОЖЕНИЕ МАТРИЦ

В алгебре нам обычно приходится выполнять две операции: сложение и умножение. До сих пор мы рассматривали только умножение матриц. Имеет ли какой-нибудь смысл сложение матриц?

Мы не встретили никаких затруднений при определении умножения. Умножение для операторов обычно означает результат последовательного выполнения операций. Применяя эту идею к матрицам, мы сразу же получили определение умножения и из него — правило умножения матриц.

Какие выгоды надеемся мы получить, перенося термины «сложение» и «умножение» из арифметики в совсем другие области?

В главе 7 мы видели, что некоторые логические задачи можно решать при помощи знаков + и ·, используя их в точности так, как если бы это были сложение и умножение в обычной арифметике или алгебре. Нам не приходилось вырабатывать новые навыки, так как наши старые навыки давали правильные результаты.

Формальная математика интересуется только закономерностями. Нас не интересуют знаки сами по себе — важны лишь закономерности, которым они подчиняются. Если некоторая идея в данной модели представляет по смыслу то же, что + в арифметике, то нет большого вреда в том, чтобы обозначать ее тем же знаком +.

Что из себя представляют сложение и умножение в арифметике? Какие допущения мы принимаем относительно знаков + и ·? К числу самых важных допущений относятся следующие:

$$a + b = b + a, \tag{I}$$
$$(a + b) + c = a + (b + c), \tag{II}$$
$$a \cdot b = b \cdot a, \tag{III}$$
$$(a \cdot b) \cdot c = a \cdot (b \cdot c), \tag{IV}$$
$$a \cdot (b + c) = a \cdot b + a \cdot c. \tag{V}$$

(I) называется законом коммутативности сложения, (III) — законом коммутативности умножения, (II) — законом ассоциативности сложения, (IV) — законом ассоциативности умножения, (V) — законом дистрибутивности.

Как правило, расставаясь с обычной арифметикой, мы опускаем требование (III). Как упоминалось раньше, мы не ожидаем, чтобы умножение непременно было коммутативным; и в самом деле, для матриц это условие обычно не выполняется.

Несколько несложных допущений следует сделать относительно операций вычитания и деления. Излагать их здесь я не собираюсь.

Сравнивая все эти допущения с аксиомами евклидовой геометрии,— особенно принимая во внимание тот факт, что Евклид использует массу терминов, не объясняя их значения,—вы увидите, что аксиомы алгебры гораздо проще геометрических аксиом. Своей простотой по сравнению с евклидовой геометрией алгебра обязана именно этому факту.

С помощью этих аксиом можно доказать любую хорошо известную алгебраическую формулу. Постараюсь объяснить формулу $(a+b)^2=a^2+2ab+b^2$ с точки зрения этих аксиом. Поскольку a^2 — сокращенная запись произведения $a\cdot a$, $2ab$ — это $a\cdot b+a\cdot b$, мне придется доказывать, что $(a+b)\cdot(a+b)=a\cdot a+(a\cdot b+a\cdot b)+b\cdot b$. Согласно аксиоме (V) $(a+b)\cdot(a+b)=(a+b)\cdot a+(a+b)\cdot b$. Чтобы перемножить $(a+b)\cdot a$, я не могу вторично воспользоваться аксиомой (V), поскольку она относится к выражениям вида $a\cdot(a+b)$, где a стоит *перед* разбиваемой на две части скобкой. Я должен обратиться к аксиоме (III), которая дает:

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

Применяя дважды аксиому (V), находим, что это равняется

$$(a\cdot a+a\cdot b)+(b\cdot a+b\cdot b)= \qquad \text{(по аксиоме (II)}$$
$$=a\cdot a+(a\cdot b+a\cdot b)+b\cdot b \qquad \text{(по аксиоме (III)}.$$

Это как раз то, что мы хотели доказать.

Возможно, вас удивляет подобный педантизм. *Для обычных чисел* это было бы действительно педантизмом. Но мы намерены применять законы алгебры к операторам любого вида, а с ними мы не так хорошо знакомы, как с арифметическими. Мы далеко не уверены, что обычные правила алгебры там применимы. Поэтому в высшей степени полезно знать, что мы имеем право пользоваться обычными формулами, если для наших операторов верны допущения (I)—(V).

Для матриц, например, мы не можем пользоваться правилом

$$(\mathbf{A}+\mathbf{B})^2=\mathbf{A}^2+2\mathbf{AB}+\mathbf{B}^2,$$

потому что в доказательстве этой формулы дважды используется аксиома (III), *которая не имеет силы для матриц.*

Мы получим право ввести в матричное исчисление знак +, если определим эту операцию так, чтобы выполнялись аксиомы (I), (II), (V). Что касается аксиомы (III), то ее выполнения мы не требуем, поскольку мы строим некоммутативную алгебру. Аксиома (IV) не содержит знака +, следовательно, на нее никак не влияет наше определение операции сложения.

Итак, можно ли определить операцию $\mathbf{U}+\mathbf{V}$ для матриц так, чтобы выполнялись свойства (I), (II), (V)? Является ли такое определение единственно возможным или же имеется несколько определений, в равной степени удовлетворяющих нашим требованиям? Последний вопрос я не собираюсь рассматривать, а только отмечу, что удовлетворительное определение операции сложения для матриц возможно. Это определение выглядит предельно просто, а именно:

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} + \begin{pmatrix} p & q \\ r & s \end{pmatrix} = \begin{pmatrix} a+p & b+q \\ c+r & d+s \end{pmatrix}.$$

Нужно просто сложить числа на одинаковых местах. Так, например:

$$\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} + \begin{pmatrix} 5 & 6 \\ 7 & 8 \end{pmatrix} = \begin{pmatrix} 1+5 & 2+6 \\ 3+7 & 4+8 \end{pmatrix} = \begin{pmatrix} 6 & 8 \\ 10 & 12 \end{pmatrix}$$

Без особого труда можно проверить, что это определение удовлетворяет всем требуемым аксиомам. Что касается аксиомы (V), то, поскольку умножение некоммутативно, стоит отметить, что оно удовлетворяет сразу двум условиям:

$$\mathbf{X}\cdot(\mathbf{U}+\mathbf{V}) = \mathbf{X}\cdot\mathbf{U} + \mathbf{X}\cdot\mathbf{V}$$

и

$$(\mathbf{U}+\mathbf{V})\cdot\mathbf{X} = \mathbf{UX} + \mathbf{VX}.$$

Иначе говоря, закон дистрибутивности выполняется независимо от того, стоит ли $\mathbf{X}$ перед или после суммы $\mathbf{U}+\mathbf{V}$.

В матричной алгебре простейшим выражением, к которому мы можем привести $(\mathbf{U}+\mathbf{V})^2$, будет результат умножения, т. е.

$$\mathbf{U}^2 + \mathbf{U}\cdot\mathbf{V} + \mathbf{V}\cdot\mathbf{U} + \mathbf{V}^2.$$

Если вы вернетесь к доказательству формулы для квадрата суммы, вы увидите, как из-за отсутствия аксиомы (III) прерывается ход доказательства.

ИМЕЕТ ЛИ СМЫСЛ В ТЕОРИИ МАТРИЦ ВЫРАЖЕНИЕ 3U+4V?

В самом начале алгебры мы встречаемся с выражениями типа $3x+4y$. Без подобных выражений алгебра не смогла бы обойтись. Можем ли мы в матричной алгебре придать какой-либо смысл выражению $3\mathbf{U}+4\mathbf{V}$?

Конечно, можем. 2U есть сокращенная запись U+U, а значение U+U уже известно из определения сложения матриц. В самом деле,

$$\mathbf{U}+\mathbf{U}=\begin{pmatrix} a & b \\ c & d \end{pmatrix}\begin{pmatrix} a & b \\ c & d \end{pmatrix}=\begin{pmatrix} 2a & 2b \\ 2c & 2d \end{pmatrix}.$$

Следовательно,

$$2\mathbf{U}=\begin{pmatrix} 2a & 2b \\ 2c & 2d \end{pmatrix}$$

Добавляя еще одно U к полученному результату. мы находим, что 3U(т. е. U + U + U) есть

$$\begin{pmatrix} 3a & 3b \\ 3c & 3d \end{pmatrix}$$

Продолжая сложение, мы видим, что для любого целого числа n

$$n\mathbf{U}=\begin{pmatrix} na & nb \\ nc & nd \end{pmatrix}$$

Как обычно, мы берем из полученного результата несколько больше, чем, строго говоря, имеем право. Эта формула была доказана для чисел n вида 2, 3, 4, 5 и т. д. Но очень заманчиво было бы предположить, с чисто формальной точки зрения, что формула эта справедлива и для чисел вроде $1\frac{3}{8}$, -7, e, $\sqrt{2}$, π и даже иногда для чисел вроде $\sqrt{-1}$. И в самом деле, оказывается, что, сделав такой шаг, мы получаем очень простую и вполне приемлемую алгебру. В соответствии с этим мы определяем в общем виде

$$k\begin{pmatrix} a & b \\ c & d \end{pmatrix}=\begin{pmatrix} ka & kb \\ kc & kd \end{pmatrix}$$

Итак, мы можем производить над матрицами следующие операции:

1) можем найти произведение UV двух матриц;
2) можем составить сумму U+V двух матриц;
3) можем умножить матрицу U на любое число k, так что произведение kU определено. (Следует отметить, что Uk означает то же самое, что и kU.)

СНОВА ПОВОРОТЫ

Имея в распоряжении эту символику, вернемся к матрице, определяющей операцию поворота:

$$\mathbf{F}=\begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$$

Глядя на эту матрицу, мы замечаем некоторую закономерность: и $\sin\theta$, и $\cos\theta$ встречаются в двух местах. Отделим эти две части друг от друга; теперь мы можем это сделать, используя формулу для суммы двух матриц. Легко видеть, что

$$\mathbf{F} = \begin{pmatrix} \cos\theta & 0 \\ 0 & \cos\theta \end{pmatrix} + \begin{pmatrix} 0 & -\sin\theta \\ \sin\theta & 0 \end{pmatrix}$$

Первая из этих матриц содержит множитель $\cos\theta$, который является числом. Мы можем вынести его за скобку, поскольку мы знаем правило умножения матрицы на число. Точно так же мы можем вынести из второй матрицы множитель $\sin\theta$. Проделав это, найдем, что

$$\mathbf{F} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \cos\theta + \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \sin\theta$$

Матрицу $\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ мы уже встречали. Она соответствует оператору I, который не меняет первоначального положения точки. Какой же буквой нам обозначать матрицу, на которую умножается $\sin\theta$? Назовем ее $\mathbf{X}$, пока не найдем для нее подходящего символа. Давайте умножим $\mathbf{X}$ на самое себя; по правилу умножения матриц получим:

$$\mathbf{X}^2 = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}$$

Но к этому же самому результату мы пришли бы, если бы умножили $\mathbf{I}$ на -1. Это можно было бы записать как $(-1)\mathbf{I}$ или, еще короче, $-\mathbf{I}$. Итак, мы имеем равенство:

$$\mathbf{X}^2 = -\mathbf{I}.$$

Поскольку $\mathbf{I}$ для матриц играет ту же роль, что 1 для чисел, это означает, что $\mathbf{X}$ можно считать в некотором смысле квадратным корнем из минус единицы. Следовательно, нашу матрицу можно обозначить буквой $\mathbf{i}$.

Если мы примем этот символ вместо $\mathbf{X}$, то равенство для $\mathbf{F}$ перепишется в виде

$$\mathbf{F} = \mathbf{I}\cos\theta + \mathbf{i}\sin\theta,$$

что сильно напоминает хорошо известное выражение

$$\cos\theta + i\sin\theta.$$

Следуя этому ходу мыслей, можно построить теорию комплексных чисел $a+ib$. Мы можем принять, что символ $a+ib$ соответствует матрице $\mathbf{I}a+\mathbf{i}b$. Написанная в развернутом виде, эта матрица выглядит следующим образом:

$$\begin{pmatrix} a & -b \\ b & a \end{pmatrix}$$

В дальнейшем будет показано, что эти матрицы при выполнении над ними операций по рассмотренным выше правилам ведут себя точно так же, как должны вести себя комплексные числа.

Было бы ошибочным считать, что подобный подход является совершенно неожиданным. Если вы вернетесь на несколько страниц назад, то перед заголовком «Применения матриц» вы увидите ту же самую матрицу

$$\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix},$$

которую мы только что использовали для оператора **i**. Там эта матрица соответствовала *операции поворота на* 90°, которая, как известно, явилась самым первым объяснением операции **i**. Матричное представление **i** дает лишь новую запись для этого геометрического представления. Будучи алгебраическим по своей природе, оно позволяет достичь более четкого изложения предмета.

МАТРИЦЫ В САМОМ ОБЩЕМ ВИДЕ

На протяжении этой главы мы рассматривали матрицы, составленные лишь из четырех чисел, расположенных в двух столбцах и двух строках. С их помощью мы могли сделать наши объяснения по возможности более простыми. Но нет причин ограничиваться рассмотрением матриц типа «два на два» или, как их называют, «2×2-матриц». Мы можем так же легко написать систему уравнений с тремя неизвестными:

$$\begin{aligned} x_1 &= ax_0 + by_0 + cz_0 \\ y_1 &= dx_0 + ey_0 + fz_0 \\ z_1 &= gx_0 + hy_0 + kz_0 \end{aligned}$$

или в матричной форме

$$\begin{pmatrix} x_1 \\ y_1 \\ z_1 \end{pmatrix} = \begin{pmatrix} a & b & c \\ d & e & f \\ g & h & k \end{pmatrix} \begin{pmatrix} x_0 \\ y_0 \\ z_0 \end{pmatrix}$$

Можно записать это и в более короткой форме, как $P_1 = \mathbf{M}P_0$. Над этими 3×3-матрицами можно производить операции сложения и умножения, причем большой разницы между их поведением и поведением 2×2-матриц нет. Правило умножения «строка на столбец» сохраняется. Квадратные матрицы в действительности могут иметь любое количество строк. Матрицы $n\times n$ практически во всем ведут себя как 2×2, за исключением их геометрической интерпретации (для матрицы $n\times n$, конечно, требуется пространство n измерений).

Можно составить и прямоугольные матрицы, где число строк не равно числу столбцов. Прямоугольные матрицы тоже можно рассматривать как «сущность» системы уравнений. Практическое применение прямоугольные матрицы находят, например, в теории электрических цепей. Прямоугольные матрицы с математической точки зрения представляют, вообще говоря, меньший интерес, чем квадратные.

Глава девятая

ОПРЕДЕЛИТЕЛИ

> *Но на следующий день один из его учеников сказал ему: «О Совершенный, зачем ты это делаешь? Хотя нам это и доставляет радость, нам не ведомы ни высокие причины, ни значение этого».*
>
> *И Савва ответил: «Сначала я вам скажу, что я делаю, а потом объясню, зачем».*
>
> Лоуренс Хоусмен,
> Совершенный

Оказывается, что знание определителей необходимо почти в любой отрасли математики. Благодаря своему широкому и разнообразному применению они привлекают к себе внимание со стороны всех, кто занимается прикладной математикой; для чистых математиков они представляют интерес как функции с особенно простыми и замечательными свойствами. Значение их очевидно, и они вполне заслуживают изучения. Правила, по которым выполняются операции над определителями, очень просты, многие задачи, касающиеся определителей, имеют зачастую удивительные и изящные решения.

Пожалуй, самым удивительным является то, что при своей простоте и очевидной математической ценности преподавание теории определителей представляет настоящую проблему. Загляните в любой учебник, где имеется раздел, посвященный определителям. Самые простые свойства объясняются с помощью ужасных вычислений. Это нельзя отнести за счет небрежности со стороны авторов. Просто не так легко найти красивый и ясный способ изложения определителей.

В этой главе я как раз собираюсь рассмотреть вопрос о том, как излагать тему «Определители». Сначала я объясняю, что такое определители, и излагаю довольно простые правила, которым они подчиняются. Если бы в моем распоряжении было только два часа, чтобы научить группу слушателей, как пользоваться определителями, мне удалось бы лишь изложить голые факты, как я это и делаю в следующем разделе. Но более способных студентов, возможно, эти голые факты не удовлетворили бы, поэтому я рассматриваю дальше вопросы, которые могли бы у них возникнуть. И, наконец, я привожу различные соображения, чтобы попытаться еще более разъяснить значение определителей. Конечно, все это не является, да и не претендует на то, чтобы быть строгим изложением вопроса. Мне бы только хотелось подвести почву под понятие определителя, показать, что теорию определителей можно изложить одновременно строго и просто, и сделать это так, чтобы свойства определителей показались естественными.

ОБЩИЕ СВЕДЕНИЯ ОБ ОПРЕДЕЛИТЕЛЯХ

Предположим, что я имею довольно большую группу слушателей и очень мало времени для того, чтобы объяснить, как использовать определители. Прежде всего, я, конечно, извинюсь за то, что недостаток времени принуждает меня изложить лишь сами факты, не объясняя их внутреннего смысла. Затем я продолжу примерно в таком духе.

Возможно, мои слушатели встречали в книгах выражение $\begin{vmatrix} a & b \\ c & d \end{vmatrix}$. Это сокращенная запись числа $ad - bc$. Если слушателям известно что-либо о матрицах или операторах или о чем-нибудь подобном, то я должен подчеркнуть, что написанное выражение — это не матрица, не оператор — это просто число. (Если бы имелась в виду матрица, то это выражение было бы заключено в круглые скобки.) Наше выражение означает просто число $ad - bc$ и ничего больше. Например,

$$\begin{vmatrix} 2 & 3 \\ 5 & 19 \end{vmatrix}$$

обозначает число 23, потому что $2 \times 19 - 3 \times 5 = 23$. Далее, выражение

$$\begin{vmatrix} a & b & c \\ d & e & f \\ g & h & k \end{vmatrix}$$

тоже является сокращенной формой записи числа

$$aek + bfg + cdh - ahf - bdk - ceg.$$

Это я также проиллюстрировал бы на числовом примере.

Квадратные таблицы чисел, ограниченные прямыми линиями с двух сторон, называются определителями. Сейчас я расскажу об их главных свойствах; всякий, кто знаком с элементарной алгеброй, может их проверить.

(I). Если поменять местами две строки определителя, то знак определителя изменится на противоположный. Например,

$$\begin{vmatrix} c & d \\ a & b \end{vmatrix} = - \begin{vmatrix} a & b \\ c & d \end{vmatrix}$$

(II). Если к числам одной строки прибавить k раз числа другой строки, то значение определителя не изменится. Например,

$$\begin{vmatrix} (a + kc) & (b + kd) \\ c & d \end{vmatrix} = \begin{vmatrix} a & b \\ c & d \end{vmatrix}$$

(III). Если поменять местами строки и столбцы, то значение определителя не изменится. Это означает, что если, например, вместо того чтобы записывать буквы a, b, c, d в алфавитном порядке по строкам, мы расположим их в алфавитном порядке по столбцам, то значение определителя не изменится:

$$\begin{vmatrix} a & b \\ c & d \end{vmatrix} = \begin{vmatrix} a & c \\ b & d \end{vmatrix}$$

Можно выразить это свойство и по-другому. Можно сказать, что отражение относительно прямой, проходящей из северо-западного угла в юго-восточный, не меняет значения определителя. Это означает, что любое предложение, справедливое в отношении строк [это касается в особенности свойств (I) и (II)], в равной мере верно и для столбцов.

(IV). Если все числа одной из строк равны нулю, то значение определителя равно нулю. Например,

$$\begin{vmatrix} a & b & c \\ 0 & 0 & 0 \\ g & h & k \end{vmatrix} = 0.$$

(V). Два определителя можно перемножить по следующему правилу (я даю здесь правило лишь для определителей 2×2):

$$\begin{vmatrix} a & b \\ c & d \end{vmatrix} \begin{vmatrix} p & q \\ r & s \end{vmatrix} = \begin{vmatrix} (ap+br) & (aq+bs) \\ (cp+dr) & (cq+ds) \end{vmatrix}$$

(VI). В заключение, мне кажется, следует указать на соотношение, которое существует между определителями 3×3 и 2×2, а именно

$$\begin{vmatrix} a & b & c \\ d & e & f \\ g & h & k \end{vmatrix} = a \begin{vmatrix} e & f \\ h & k \end{vmatrix} - b \begin{vmatrix} d & f \\ g & k \end{vmatrix} + c \begin{vmatrix} d & e \\ g & h \end{vmatrix}$$

Определитель 2×2, стоящий множителем при a, получился из определителя 3×3 вычеркиванием строки и столбца, содержащих a; остается

$$\begin{matrix} e & f \\ h & k \end{matrix}$$

Два других определителя, стоящих в правой части уравнения, получены тем же способом: один из них — вычеркиванием строки и столбца, содержащих b, а другой — вычеркиванием строки и столбца, содержащих c.

Обобщение этого правила позволяет нам определить определители 4×4. С помощью соотношения

$$\begin{vmatrix} a & b & c & d \\ e & f & g & h \\ j & k & m & n \\ p & q & r & s \end{vmatrix} = a \begin{vmatrix} f & g & h \\ k & m & n \\ q & r & s \end{vmatrix} - b \begin{vmatrix} e & g & h \\ j & m & n \\ p & r & s \end{vmatrix} + c \begin{vmatrix} e & f & h \\ j & k & n \\ p & q & s \end{vmatrix} - d \begin{vmatrix} e & f & g \\ j & k & m \\ p & q & r \end{vmatrix}$$

мы вычисляем определитель 4×4, стоящий в левой части равенства.

А значения четырех определителей 3×3, стоящих справа, нам известны, поскольку определители 3×3 уже введены.

Надеюсь, уже ясно, что, повторив подобную процедуру, можно получить определитель 5×5, а из него 6×6 и так до бесконечности. И в каждом случае мы смогли бы проверить, при условии, что у вас хватит энергии проделать все необходимые вычисления, что свойства (I), (II), (III), (IV) и (V) выполняются и для этих более громоздких определителей.

Шесть правил — это не так уж много для прилежного студента. Правила, как видите, не особенно сложные. За два академических часа, в промежутке между которыми учащиеся немного поупражняются самостоятельно дома, можно, вероятно, дать им полное представление о том, что можно делать с определителями. Нужно проработать несколько примеров на доске и проверить, как учащиеся выполняют сами такого же типа примеры.

КРИТИКА МЕТОДА

Какие критические замечания мог бы сделать умный студент относительно подобного изложения материала?

Первое замечание, вероятно, касалось бы того, что данные определения несколько произвольны. Я начинаю с определения определителей 2×2 и 3×3 и с помощью (VI) распространяю это определение на определители с большим числом членов. Но как я прихожу к этим определениям? Почему я считаю естественным начать именно с таких определений?

Второе замечание относилось бы к методу доказательства. Я предлагаю студентам проверить с помощью алгебраических вычислений, что все утверждения, сделанные мною относительно определителей 2×2, 3×3, 4×4 и 5×5, верны. Но, очевидно, я уверен, что свойства (I)—(V) в равной степени верны для 6×6 и 7×7 и вообще для любых $n\times n$-определителей. Откуда же взялась у меня уверенность, что определители любого порядка, вычисленные согласно данному мною определению, будут обладать теми же самыми свойствами?

Третье замечание касалось бы вопроса изящества доказательств. Свойства (I) — (V) очень просты. Но методы, которые я предлагаю для их проверки, предполагают выполнение длинных и громоздких вычислений. Если результаты столь просты, нет ли более простого способа их доказательства? Если я действительно понимаю, как возникли эти результаты, то почему бы мне не доказать их способом, содержащим гораздо меньше вычислений, или даже вообще без вычислений?

Конечно, если обучение происходит в спешке, подобную критику можно отмести как неуместную. Но в том случае, когда имеется хоть какая-нибудь возможность остановиться и вникнуть в суть дела, эти критические замечания вполне оправданы.

Мне бы не хотелось, чтобы остаток этой главы воспринимался как предписание совершенного метода преподавания определителей. Конечно, я не сомневаюсь в том, что, имей я больше свободного времени и возможностей, где-нибудь в библиотеках мира я бы нашел готовое и полное, гораздо лучшее изложение этого вопроса. То, что я пишу здесь, следует рассматривать лишь как вехи того пути, по которому можно следовать, чтобы искать способы лучшего изложения идеи определителей.

ОТ МАТРИЦ К ОПРЕДЕЛИТЕЛЯМ

На рис. 44 (стр. 95) показан результат применения матричной операции. Мы видим, что квадратики преобразуются в параллелограммы. Рассмотрим действие этой операции на величину площади любой фигуры. Если первоначально фигура состояла из некоторого числа квадратиков, то после операции она будет состоять из такого же числа параллелограммов. Доли квадрата будут преобразованы в такие же доли параллелограмма. Это означает, что все площади изменяются в одном и том же отношении. Если площадь параллелограмма в g раз больше площади квадрата, то все площади после операции будут в g раз больше, чем до операции.

Чему равняется число g для матрицы

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}?$$

Напомним, что вершины параллелограмма, отличные от начала координат O, имели координаты (a, c), (b, d) и $(a+b;\ c+d)$. На рис. 45 этот параллелограмм изображен заключенным в прямоугольник. Указаны длины различных горизонтальных и вертикальных участков. Площадь параллелограмма может быть найдена исключением из площади прямоугольника площадей всех незаштрихованных участков, а именно двух прямоугольников площадью bc каждый, двух треугольников с площадями $\frac{1}{2}ac$ и двух треугольников с площадями $\frac{1}{2}bd$. Поскольку площадь большого прямоугольника равна $(a+b)\ (c+d)$, то площадь параллелограмма оказывается равной $(a+b)(c+d)-(2bc-ac-bd)$, или, после упрощения, $ad-bc$. Квадрат, из которого был получен наш параллелограмм, имел вершины в точках с координатами $(0,0)$, $(0,1)$, $(1,0)$ и $(1,1)$. Следовательно, площадь квадрата равна 1, откуда

$$g=ad-bc.$$

Мы узнаем ответ. Это ведь не что иное, как определитель

$$\begin{vmatrix} a & b \\ c & d \end{vmatrix}$$

Итак, у нас имеется теперь интерпретация для определителя 2×2: *он представляет собой отношение, в котором матрица меняет площади.*

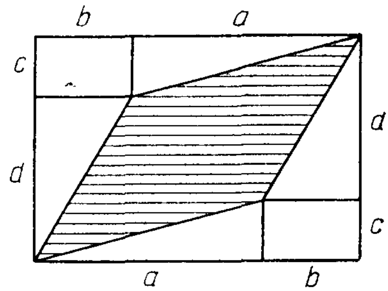


Рис. 45.

Эта идея годится не только для определителей 2×2. Если мы рассмотрим матрицы 3×3, то увидим, что они подобным же образом воздействуют на точки (x, y, z) трехмерного пространства. Каждая матрица преобразует куб в косоугольный параллелепипед. Далее, матрица изменяет объем пространственных тел в постоянном отношении, причем это отношение и является определителем матрицы. Мы не можем себе наглядно представить соответствующий процесс в пространстве четырех или более измерений, но логически здесь нет никакой разницы. Обнаруженная нами идея применима в принципе ко всем $n\times n$-матрицам, для любого n.

Следовательно, определитель — это число, связанное с матрицей, вот почему мы записываем их так, что одно нам напоминает другое.

Во всех случаях, когда для сокращенной записи матрицы используем букву **U**, соответствующий ей определитель мы будем обозначать $|\mathbf{U}|$; этот символ является сокращенной записью слов «определитель матрицы **U**».

УМНОЖЕНИЕ ОПРЕДЕЛИТЕЛЕЙ

Возможно, вы уже обратили внимание на то, что в свойстве (V) определителей используется правило, знакомое нам по главе 8 как правило умножения матриц. Если вы вернетесь к разделу «Умножение матриц» главы 8, то увидите, что правило умножения получилось очень просто и естественно. Его не пришлось высасывать из пальца, оно естественно возникло в процессе работы.

Теперь легко понять, почему это же самое правило позволяет нам найти произведение двух определителей. Положим, мы имеем две матрицы, **U** и **V**. Пусть $|\mathbf{U}|=g$, $|\mathbf{V}|=h$. Это означает, что при выполнении операции **U** любая площадь увеличивается в g раз, а применение операции **V** увеличивает каждую площадь в h раз. Произведение **UV** означает просто результат применения сначала операции **V**, а затем **U**. Чему будет равен определитель матрицы **UV**? Другими словами, в каком отношении матрица **UV** изменяет площадь фигур? Ответ очевиден. Операция **U** увеличивает площадь в h раз, затем операция **U** увеличивает площадь в g раз. Ясно, что совместный эффект двух операций сказывается в увеличении площади в gh раз. Поэтому $|\mathbf{UV}|=gh$, т. е. если мы перемножаем две матрицы по правилу умножения матриц и находим определитель результата, мы получаем произведение определителей обеих матриц. Это можно записать в виде

$$|\mathbf{UV}| = |\mathbf{U}|\ |\mathbf{V}|.$$

Важность этого результата заключается не в его применении к вычислению произведения двух определителей, потому что если нам известна величина g одного определителя и величина h другого, то гораздо легче перемножить два числа g и h, чем прибегать к правилу (V). Вероятно, лучше использовать это правило в обратном направлении. Предположим, что нам нужно подсчитать сложный определитель. Мы замечаем, что его числа составлены так же, как в произведении матриц **UV** [т. е. как в правой части равенства, выражающего правило (V)], и можем тогда свести задачу к вычислению двух более простых определителей: $|\mathbf{U}|$ и $|\mathrm{V}|$. Этот факт можно также использовать для обоснования правил (I) и (II), что и будет нами сделано в одном из следующих разделов этой главы.

РАСТЯЖЕНИЯ, ПОВОРОТЫ И ОТРАЖЕНИЯ

В главе восьмой мы нашли матрицы, представляющие различные геометрические операции. Действие этих операций на площади фигур легко усматривается из геометрических соображений. Если мы обратимся теперь к определителям рассматриваемых матриц, то мы не только сможем проверить ряд положений, но и получим один неожиданный результат. Определитель матрицы **G** (глава 8) есть

$$\begin{vmatrix} k & 0 \\ 0 & k \end{vmatrix} = k\cdot k - 0\cdot 0 = k^2.$$

Но это и понятно: поскольку **G** увеличивает длины в k раз, совершенно очевидно, что площади увеличиваются при этом в k^2 раз [1].

Операция поворота лишь перемещает площадь без изменения ее величины. Следует ожидать поэтому, что значение определителя для операции поворота окажется равным 1. И действительно, для операции C, представляющей поворот на 180°, имеем:

$$|\mathbf{C}| = \begin{vmatrix} -1 & 0 \\ 0 & -1 \end{vmatrix} = 1.$$

[1] Здесь и всюду я использую термин, применимый, строго говоря, только в определенных условиях. **G** представляет собой растяжение только в том случае, когда k больше 1. Если k — правильная дробь (<1), то **G** означает сжатие. Если учесть, что возможны и отрицательные значения k, то мне пришлось бы ввести дальнейшие поправки в рассуждения. Но я считаю, что с точки зрения экономии умственных усилий гораздо легче рассматривать сначала самые простые, самые очевидные ситуации. Впоследствии можно будет проверить, применимы ли эти результаты и для других случаев.

Для операции **F**, которая представляет собой поворот на угол θ, имеем:

$$|\mathbf{F}| = \begin{vmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{vmatrix} = (\cos\theta)(\cos\theta) - (\sin\theta)(\sin\theta) = \cos^2\theta + \sin^2\theta = 1.$$

И это тоже вполне понятно.

Остаются еще операции отражения **A**, **B**, **E**, **D**. Для них мы находим (после упрощений, подобных тем, что использовались при вычислении $|\mathbf{F}|$):

$$|\mathbf{A}| = \begin{vmatrix} 1 & 0 \\ 0 & -1 \end{vmatrix} = -1,$$

$$|\mathbf{D}| = \begin{vmatrix} \cos 2\alpha & \sin 2\alpha \\ \sin 2\alpha & -\cos 2\alpha \end{vmatrix} = -1,$$

$$|\mathbf{B}| = \begin{vmatrix} -1 & 0 \\ 0 & 1 \end{vmatrix} = -1, \quad |\mathbf{E}| = \begin{vmatrix} 0 & 1 \\ 1 & 0 \end{vmatrix} = -1.$$

Это уже нечто неожиданное. Нам кажется естественным, что при повороте куска картона некоторой формы его площадь не меняется; однако вычисление показывает нам, что площадь умножается на -1.

В повседневной жизни мы представляем себе площадь всегда положительной величиной. Но когда мы начинаем ее подсчитывать, она может оказаться отрицательной. Так, из курса дифференциального и интегрального исчислений хорошо известно, что при нахождении площади интегрированием результат может быть отрицательным. Или по известной формуле школьной математики площадь треугольника равна

$$\sqrt{s(s-a)(s-b)(s-c)}.$$

Здесь мы имеем квадратный корень, а при извлечении квадратного корня всегда возникает вопрос о знаке. Следовательно, если рассматривать вопрос чисто алгебраически, треугольник имеет две площади: одну положительную и другую отрицательную. Площадь в повседневной жизни — это просто величина, без учета знака.

Однако даже в повседневной жизни встречаются случаи, когда наличие у площади двух знаков имеет некоторый смысл. К паровому двигателю прикрепляется устройство, известное под названием индикатора. Карандаш связан с измерителем давления, а кусок бумаги — с некоторыми движущимися частями машины так, чтобы в процессе работы двигателя на бумаге получилась некоторая диаграмма. Диаграмма имеет вид замкнутой кривой; площадь, охваченная этой кривой, показывает, какую работу проделывает тепловой двигатель при каждом ходе. А теперь, вместо того чтобы представлять двигатель работающим, предположим, что мы сами производим работу над ним. Мы будем толкать его так, чтобы он двигался в обратном направлении; карандаш пойдет по уже начерченной

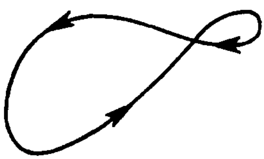


Рис. 46.

кривой, но в противоположном направлении. Площадь кривой представляет теперь не ту работу, которую произвел двигатель, а ту, которую произвели над ним, т. е. *отрицательную работу*. Поэтому ту же самую кривую, нарисованную в направлении, противоположном нормальному, следует рассматривать как имеющую отрицательную площадь.

Если кому-нибудь удалось бы построить двигатель, который описывал бы кривую в виде восьмерки (вроде кривой, показанной на рис. 46), это означало бы, что на протяжении одной части цикла, скажем, большей петли, которую он рисует в направлении против часовой стрелки, он проделывал бы полезную работу, а в меньшей петле, которая описывается по часовой стрелке, он поглощал бы энергию.

Если вам не по вкусу идея отрицательных площадей, объемов и т. п., мы могли бы пойти вам навстречу и перефразировать наше утверждение относительно связи между определителями и площадями. Мы могли бы сказать, что определитель матричной операции указывает на две вещи: величина определителя указывает на отношение, в котором меняются площади в результате операции, а знак говорит о том, имеет ли место изменение ориентации. Например, в трехмерном пространстве матричная операция, определитель которой отрицателен, превратила бы автомобиль, у которого руль справа, в автомобиль с рулем с левой стороны, т. е. такая операция давала бы зеркальное изображение мира.

Но сама идея умножения площадей или объемов на -1 оказывается иногда чрезвычайно удобной. Например, мы знаем, что никаким числом поворотов нам не удалось бы превратить автомобиль с рулем справа в автомобиль с рулем слева; иными словами, путем комбинации вращений невозможно получить отражение. Могли бы мы доказать это математически? Да, но при одном предположении: если бы нам было позволено говорить об отрицательных и положительных объемах. Любое вращение умножает объемы на $+1$; если последовательно произвести несколько вращений, то, поскольку каждое из них умножает объем на $+1$, в итоге мы никак не придем к умножению исходного объема на -1. Только операция отражения позволяет нам сделать это.

ВЫВОД СВОЙСТВ (I) И (II)

Теперь представьте себе следующую несложную процедуру. У нас имеется матрица **U**, состоящая из двух строк и двух столбцов, ее определитель равен g. Мы выполняем сначала операцию **U**, а затем операцию **E**.

Каково будет результирующее воздействие этих двух операций на площади плоских фигур? Операция **U** умножает площади на g. Операция **E**, являющаяся отражением относительно прямой $y=x$, как мы видели в предыдущем разделе, умножает площади на -1. Суммарным эффектом оказывается умножение площади на $(-1)g$, т. е. на $-g$. Следовательно, определитель сложной операции **EU** должен быть равен $-g$:

$$|\mathbf{EU}| = -g.$$

Что же в действительности дает нам операция **EU**?

Положим $\mathbf{U} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$; тогда, поскольку $\mathbf{E} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$,

$$\mathbf{EU} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} c & d \\ a & b \end{pmatrix}$$

Таким образом, матрица **EU** получается из матрицы **U** перестановкой ее строк. А ее определитель, как мы видели выше, равен $-g$, т. е. определителю исходной матрицы, взятому с противоположным знаком. Так мы доказали правило (I).

Если вы проделаете то же самое относительно операции **UE**, вы получите соответствующее правило для перестановки столбцов.

Правило (II) может быть получено таким же образом, т. е. путем последовательного выполнения двух операций.

Пусть матрица **H** представляет операцию, которая превращает единичный квадрат в параллелограмм, показанный ниже на рисунке 47.

Эта операция не предполагает никакого изменения площади. Треугольник, на который увеличивается площадь фигуры справа, в точности компенсирует то, что фигура теряет слева. Нет здесь и отражения. Превращение квадрата в параллелограмм может быть выполнено постепенно, как если бы кто-нибудь упирался в верхний левый угол квадрата со все возрастающей силой; между тем отражение никогда нельзя выполнить постепенно, если речь идет о плоскости. Следовательно, определитель **H** должен быть равен $+1$. Отсюда

$$|\mathbf{HU}| = |\mathbf{H}|\ |\mathbf{U}| = 1g = g.$$

Мы видим, что матрицы **HU** и **U** имеют один и тот же определитель.

Что такое **HU**? Прежде всего, нам необходимо знать все четыре члена матрицы **H**. Их нетрудно найти, пользуясь методом, объясненным в конце раздела «Матрица в общем виде» главы 8. Операция **H** переводит точку (1,0) в положение (1,0); следовательно, первый столбец должен быть

$$\begin{matrix} 1 \\ 0 \end{matrix}$$

Далее, **H** переводит точку (0,1) в положение (k, 1); следовательно, второй столбец должен быть

$$\begin{matrix} k \\ 1 \end{matrix}$$

(Информацию о положении точек мы берем из рис. 47.) Поэтому операция **H** должна иметь вид:

$$\begin{pmatrix} 1 & k \\ 0 & 1 \end{pmatrix}$$

и, следовательно,

$$\mathbf{HU} = \begin{pmatrix} 1 & k \\ 0 & 1 \end{pmatrix}\begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} a + kc & b + kd \\ c & d \end{pmatrix}$$

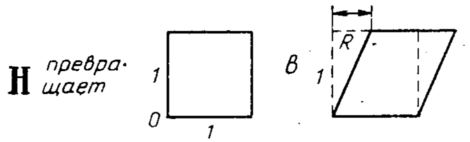


Рис. 47.

Но мы уже видели, что определитель матрицы **HU** равен определителю матрицы **U**. Как раз это и гласит правило (II). Матрица

$$\begin{pmatrix} 1 & 0 \\ k & 1 \end{pmatrix}$$

имеет примерно тот же геометрический смысл, что и матрица **H**: она тоже преобразует квадрат в равновеликий ему параллелограмм, только смещение происходит в вертикальном направлении, а не в горизонтальном. С ее помощью мы можем доказать, что

$$\begin{vmatrix} a & b \\ c + ka & d + kb \end{vmatrix} = \begin{vmatrix} a & b \\ c & d \end{vmatrix}$$

т. е. мы можем k раз прибавлять первую строку ко второй.

Такое же правило можно получить и для столбцов, если рассматривать произведения в обратном порядке, например **UH**.

ОБОБЩЕНИЕ МЕТОДА

Что из себя представляет матрица 3×3, соответствующая операции **E**? Это легко видеть из рассмотрения уравнений, соответствующих $P_1 = \mathbf{E}P_0$:

$$x_1 = y_0,$$
$$y_1 = x_0.$$

Мы замечаем здесь перестановку двух букв. Новое x равно старому y, а новое y равно старому x.

Если мы имеем дело с тремя переменными x, y, z, то пользуемся той же самой идеей, а именно переставляем любые две из трех букв. Третья буква, конечно, останется на своем месте. Например, поменяем местами x и y, а величину z оставим такой, как она была; тогда мы получим:

$$\begin{aligned} x_1 &= \quad y_0, \\ y_1 &= x_0, \\ z_1 &= \qquad z_0, \end{aligned}$$

что записывается с помощью матрицы

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

Если бы мы переставили y и z, то получили бы матрицу

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix},$$

а перестановка x и z дает матрицу

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

Если вы перемножите какую угодно 3×3-матрицу слева на одну из этих матриц, вы увидите, что в результате две строки поменяются местами.

Геометрически каждая из этих матриц представляет эффект отражения в плоском зеркале. Если считать, что x, y, z обозначают соответственно «восток», «север», «вверх», то первая матрица будет соответствовать отражению в вертикальном зеркале, которое расположено в северо-восточном направлении. Это показано на рис. 48.

Таким же путем мы можем найти матрицу, которая в пространстве трех измерений соответствует операции **H**. Эта матрица имеет вид:

$$\begin{pmatrix} 1 & k & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

Получается так, как если бы куб слегка толкнули и две его противоположные грани оказались в наклонном положении, а объем остался без изменения. Можно представить себе кусок сыра, у которого с одной стороны отрезали под углом ломтик и приставили его к другой стороне.

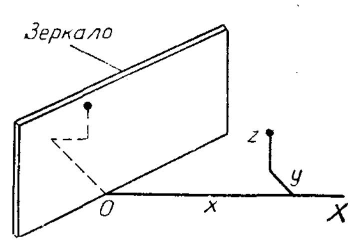


Рис. 48.

В действительности, этот метод можно применить для любых $n\times n$-матриц. Трудность заключается лишь в том, что для отыскания строгого изложения нам пришлось бы предварительно объяснять, что имеется в виду под словом «объем» в пространстве n измерений. Мы не собираемся здесь вдаваться в эти детали. Для нас совершенно достаточно было показать, что определители действительно имеют простой геометрический смысл и что их свойства, во всяком случае свойства определителей 2×2 и 3×3, легко могут быть доказаны геометрически.

Стоило бы, пожалуй, указать здесь на один пункт, где этот метод не действует: он не устанавливает свойства (III) определителей. Дело в том, что свойство (III) существенно отличается от свойств (I) и (II). Невозможно найти такие матрицы, которые бы годились для доказательства свойства (III), подобно тому как мы нашли матрицы **E** и **H** для свойств (I) и (II). Другими словами, невозможно найти матрицы, составленные из постоянных чисел, которые с помощью умножения превращали бы

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \text{ в } \begin{pmatrix} a & c \\ b & d \end{pmatrix}$$

Мы не будем здесь обсуждать, как заполнить этот пробел.

ОСОБЕННЫЕ МАТРИЦЫ

Если мы посмотрим на матрицу

$$\mathbf{M} = \begin{pmatrix} 1 & 1 \\ 2 & 2 \end{pmatrix},$$

то увидим, что ее определитель равен нулю. Таким образом, при действии этой матрицы любая площадь умножается на 0.

Выясним, что происходит с квадратом $OABC$ на рис. 49 в результате применения операции **M**. Здесь O — начало координат, A — точка (1,1), B — точка (0,2), C — точка (—1,1).

Запишем уравнения, соответствующие операции **M**:

$$x_1 = x_0 + y_0,$$
$$y_1 = 2x_0 + 2y_0.$$

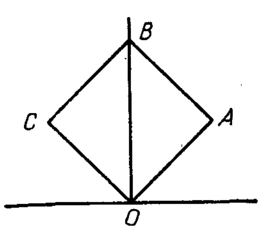


Рис. 49.

Подставляя в эти уравнения вместо x и y координаты точек O, A, B, C, получаем, что начало координат (0, 0) переходит в (0, 0), точка (1, 1) переходит в (2, 4), точка (0,2) — в (2,4), точка (—1,1) — в (0,0). Итак, точки O и C обе переходят в положение (0,0), а точки A и B—в положение (2,4).

Другими словами, матрица **M** сжимает стороны OC и AB до предела, превращая их в нуль. Именно поэтому площадь квадрата становится тоже равной нулю. Но площадь любого другого квадрата также становится равной нулю. Вы можете убедиться, что какие бы значения вы ни придавали x_0 и y_0, точка с координатами (x, y) всегда лежит на прямой $y=2x$. Таким образом, матрица сжимает любую плоскую фигуру в линию.

Матрица, определитель которой равен нулю, называется особенной. Особенная матрица всегда переводит некоторую точку, вроде точки C в нашем примере, в начало координат.

Оба эти условия в действительности эквивалентны. Если некоторая точка не лежит в начале координат, но в результате матричной операции она туда попадает, то определитель этой матрицы равен нулю. И обратно, если определитель матрицы равен нулю, то в результате операции некоторая точка, отличная от начала координат, непременно окажется в начале координат.

Это означает, что определитель матрицы

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

равен нулю тогда и только тогда, когда имеется некоторая точка (x, y), отличная от начала координат и такая, что $(ax+by, cx+dy)$ есть $(0,0)$, т. е. уравнения

$$ax+by=0,$$
$$cx+dy=0$$

имеют решение, отличное от $x=0$, $y=0$.

Последнее утверждение можно обобщить на матрицы 3×3 и вообще на любые квадратные матрицы. Вероятно, чаще всего определители встречаются в алгебре именно в связи с этим условием.

АЛГЕБРАИЧЕСКИЙ ПОДХОД К ОПРЕДЕЛИТЕЛЯМ

Если иметь в виду приведенное выше условие равенства определителя нулю, то нетрудно понять, сколь «разумно» то, что определители имеют перечисленные ранее свойства.

Рассмотрим, например, связь между определителями

$$\begin{vmatrix} a & b \\ c & d \end{vmatrix} \text{ и } \begin{vmatrix} c & d \\ a & b \end{vmatrix}$$

Если мы обозначим их соответственно через D_1 и D_2, то, как мы знаем,

$$D_2=-D_1;$$

но это есть просто правило (I). Могли бы мы предвидеть это соотношение?

Рассмотрим значения обоих определителей. $D_1=0$ является условием того, что пара чисел x и y, отличных от нуля, удовлетворяет уравнениям

$$ax+by=0, \tag{1}$$

$$cx+dy=0. \tag{2}$$

Точно так же $D_2=0$ является условием того, что два отличных от нуля числа удовлетворяют уравнениям

$$cx+dy=0, \tag{3}$$

$$ax+by=0. \tag{4}$$

(Уравнения (3) и (4) написаны в соответствии с расположением букв во втором определителе.)

Но уравнения (3) и (4) — это просто уравнения (1) и (2), написанные в обратном порядке. Естественно, что если существует решение x, y, отличное от 0, 0, когда уравнение $ax+by=0$ написано перед уравнением $cx+dy=0$, то такое же решение должно существовать и в том случае, когда $ax+by=0$ записано после уравнения $cx+dy=0$. Другими словами, если $D_1=0$, то и D_2 должен быть равен 0. И обратно, если уравнения (3) и (4) имеют решение, то имеют решение и уравнения (1) и (2), т. е. $D_1=0$. Таким образом, определители D_1 и D_2 тесно связаны между собой. Если один из них равен нулю, то равен нулю и другой.

Приложив некоторое старание, можно было бы вывести из этого рассмотрения (для определителей с любым числом строк и столбцов) правило (I). В процессе доказательства встретятся кое-какие затруднения; поэтому если вам и не хочется подобным образом проверять правило (I), то вы, по крайней мере, можете убедиться, что оно вполне естественно и разумно.

Аналогично, правило (II) соответствует тому факту, что к уравнению (1) можно добавлять k раз уравнение (2):

$$(a+kc)x+(b+kd)y=0, \tag{5}$$

$$cx+dy=0. \tag{6}$$

Уравнения (5) и (6) означают в точности то же самое, что уравнения (1) и (2), поэтому не удивительно, что определители в этих двух случаях равны.

Поменять местами столбцы в определителе — это значит записать члены, содержащие y, впереди членов, содержащих x:

$$by+ax=0, \tag{7}$$

$$dy+cx=0. \tag{8}$$

При такой записи меняется знак соответствующего определителя. Этого вполне можно было ожидать, поскольку тривиальная перестройка уравнений должна как-то просто сказываться на определителе.

Столь же просто мы могли бы показать, что и все остальные свойства определителей соответствуют очевидным свойствам соответствующих систем уравнений. Нам удалось избежать громоздких вычислений, с которых мы начали эту главу, и перейти в сферу идей. По-видимому, алгебраический подход позволяет дать более простые доказательства, чем геометрический, при помощи матриц; нам здесь не нужно задаваться вопросом, что означает объем в пространстве n измерений.

КВАДРАТНЫЕ УРАВНЕНИЯ С ТРЕМЯ КОРНЯМИ

В алгебре хорошо известно, что уравнение второй степени не может иметь три различных корня; если три числа a, b и c удовлетворяют квадратному уравнению, то два из них должны совпадать: $a=b$, или $b=c$, или $c=a$. Давайте посмотрим, как этот хорошо известный результат выражается через определители.

Мы ищем такое квадратное уравнение $px^2+qx+r=0$, которому бы удовлетворяли три числа a, b и c. Подставляя a, b и c в уравнение, мы находим, что они будут удовлетворять этому уравнению, если

$$\begin{aligned} a^2p + aq + r &= 0, \\ b^2p + bq + r &= 0, \\ c^2p + cq + r &= 0. \end{aligned}$$

Но для этих равенств имеется и матричная форма записи. Мы можем сказать, что матрица

$$\begin{pmatrix} a^2 & a & 1 \\ b^2 & b & 1 \\ c^2 & c & 1 \end{pmatrix}$$

переводит точку (p, q, r) в начало координат $(0, 0, 0)$.

Здесь имеются две возможности. Возможно, что точка (p, q, r) уже расположена в начале координат, тогда матрица ничего необычного не делает. Поскольку в этом случае $p=0$, $q=0$ и $r=0$, то искомое уравнение, которому должны удовлетворять наши числа a, b и c, будет иметь вид: $0x^2+0x+0=0$, что не слишком-то похоже на уравнение. Любое число может быть его корнем.

Но если точка (p, q, r) не лежит в начале координат, тогда матрица должна быть особенной. Поэтому, для того чтобы числа (a, b и c) были корнями настоящего квадратного уравнения (уравнения, отличного от $0x^2+0x+0=0$), должно выполняться следующее условие:

$$\begin{vmatrix} a^2 & a & 1 \\ b^2 & b & 1 \\ c^2 & c & 1 \end{vmatrix} = 0$$

Во многих задачниках по алгебре имеется упражнение, где нужно показать, что этот определитель равен $(a-b)$ $(a-c)$ $(b-c)$. Тот факт, что этот определитель можно записать в виде произведения трех множителей, не удивителен: как мы уже отмечали в начале раздела, для того чтобы a, b и c являлись корнями (настоящего) квадратного уравнения, должно быть $a=b$ или $b=c$ или $c=a$, что соответствует трем сомножителям данного произведения.

ОПРЕДЕЛИТЕЛИ КАК ИСТОЧНИК ЗАКОНОМЕРНОСТЕЙ

Определители часто помогают придать форму тому, что без них было бы бесформенным. Представьте, например, что вам нужно обобщить хорошо известный элементарный результат

$$a^2-b^2=(a+b)(a-b).$$

На первый взгляд выражение a^2-b^2 не содержит никакой закономерности. Но вам может прийти в голову записать его в виде определителя

$$\begin{vmatrix} a & b \\ b & a \end{vmatrix}$$

Здесь начинает вырисовываться некоторая закономерность, и вы можете далее рассмотреть определитель

$$\begin{vmatrix} a & b & c \\ b & c & a \\ c & a & b \end{vmatrix}$$

Произведя подсчет, получим $a^3+b^3+c^3-3abc$. Это выражение содержит множитель $(a+b+c)$. Очевидно, что в случае 4, 5 и вообще любого числа членов можно поступать подобным же образом.

В учебниках алгебры студенты часто встречаются с выражением $a^3+b^3+c^3-3abc$. Возможно, им непонятно, почему именно это выражение встречается так часто. Определители же показывают, что это выражение отражает некоторую закономерность и что оно должно обладать рядом простых свойств. Это обстоятельство делает его особенно подходящим в качестве материала для упражнений.

Глава десятая

ПРОЕКТИВНАЯ ГЕОМЕТРИЯ

Бесконечность — это место, где происходит то, чего не бывает.

Из ответа школьника

Проективная геометрия — один из самых красивых разделов математики.

Несомненно, что для профессионального математика она составляет существенную часть специального образования. Вовсе не обязательно изучать ее досконально, это имеет смысл лишь для тех, кто будет специализироваться в данной области; однако основные закономерности проективной геометрии можно проследить во многих других отраслях математики, они помогают направить исследование и унифицировать разрозненные результаты.

Нематематикам также стоит изучать этот раздел. Я говорю здесь не о его технической ценности. Правда, проективная геометрия имеет некоторую связь с теорией аэрофотосъемки, но не это заставляет меня высказываться в пользу включения проективной геометрии в программу обучения. Ее ценность скорее заключается в том, что она оживляет курс. Нематематик, изучая математику для технических целей, часто вынужден осиливать массу шаблонных приемов, которые бывают чрезвычайно скучными. Это притупляет остроту умственного восприятия. Программа обучения должна включать такие элементы, которые бы действовали наподобие холодного душа — они должны производить встряску и поддерживать бодрое состояние ума.

Проективная геометрия наилучшим образом служит этим целям. Она удивительна: она делает, казалось бы, непозволительные вещи. Она изобилует чудесными невозможностями. Параллельные линии там пересекаются, и имеется даже одна теорема (совершенно достаточная для того, чтобы заставить нормального человека усомниться в здравом смысле математиков), где утверждается, что все окружности имеют две общие точки. Это, конечно, не обычные точки, они воображаемые и находятся в бесконечности. И все-таки — результат достаточно поразительный.

Более того, прективная геометрия является превосходным примером математического стиля. Если в ней что-нибудь можно доказать, то это доказывается обычно очень просто. В этом отношении она представляет полную противоположность евклидовой геометрии. В главе второй (на примере решения задачи о воде и вине) мы встретились с принципом, что вопрос становится одновременно общим и простым, если *отбросить всю ненужную информацию.* Именно это и делает проективная геометрия. Она упорно игнорирует определенный тип информации, который часто является причиной сложности евклидовой геометрии.

К тому же, ее история чрезвычайно интересна и показывает, как мало мы обращаем внимания на возможные последствия всего, что мы делаем. Проективная геометрия возникла как очень практический предмет, а именно как теория перспективы. Если художнику нужно нарисовать стол или коробку, как он должен это делать? Бывает, что художник оказывается инженером; и вот как раз Г. Дезарг (1593—1662), которому проективная геометрия обязана своим рождением, был одновременно и архитектором и инженером.

Теория перспективы очень интересна сама по себе. Она, конечно, исходит из обычной евклидовой геометрии и применяет ее к теории построения рисунка. Проективная геометрия возникла как один из разделов евклидовой геометрии. Понадобились века для того, чтобы люди поняли, что она является на самом деле самостоятельным предметом, что она может быть создана без всякого обращения к евклидовой геометрии и что лучший способ очистить евклидову геометрию (которая включает в себя множество несформулированных предположений) заключается в том, чтобы создать сначала проективную геометрию, а уже на ее основе — евклидову. Проективная геометрия сегодня — это стройный, ясный, логичный предмет, каким никогда не была евклидова.

В связи с этим перед преподавателем возникает вопрос, должен ли он придерживаться исторического подхода. В пользу такого подхода говорит тот факт, что если ученики увидят и поймут, как возникла проективная геометрия, то они воспримут это как естественное развитие. Против исторического подхода имеется одно важное возражение: евклидова геометрия нелогична, проективная геометрия логична. Обидно делать логичную вещь зависимой от нелогичной! Как обидно студентам учить то, что они потом будут переучивать! С другой стороны, если начать изучение проективной геометрии исходя из современной точки зрения, все будет ясно и логично, однако учащиеся не получат представления об идее предмета, а также о том, откуда он возник.

Мне кажется, надо начать с того, что показать студентам историческое развитие, но с самого же начала предупредить их, что предмет достиг сейчас такой стадии развития, с которой он должен был бы в действительности начинаться. С логической точки зрения (для ангела по телефону) современное развитие предмета несомненно выше: количество допущений незначительно, и они весьма просты, нет необходимости прибегать к чертежам или физическим представлениям о формах и размерах. Но как раз потому, что этого совершенно достаточно для логиков и ангелов, для большинства людей этого недостаточно. Гораздо поучительнее видеть, как ясные, логичные идеи постепенно вырисовываются из беспорядочного нагромождения понятий, чем просто начинать с изложения этих идей. Если четко объяснить, в каких пределах применимы старые методы, то, мне думается, не так уж много вещей придется «переучивать». Студент в состоянии почувствовать разницу между «доказательством» 1640 года и доказательством, которое стало приемлемым после 1899 года (год, когда Гильберт опубликовал свои «Основания геометрии»). Не знаю, сочтут ли необходимым математики будущих веков говорить о «так называемых доказательствах», принятых математиками XX века,— возможно, что да. Но, во всяком случае, по сравнению с XVII веком мы сделали большой шаг вперед.

Несколько раньше я уже упоминал, что практическая ценность проективной геометрии заключается не в ее непосредственном применении для аэрофотосъемки или для черчения, а в том влиянии, которое она оказывает на другие области математики. Можно привести немало примеров этого влияния; в некоторых случаях потребовались бы длинные объяснения. В данный момент мы остановимся лишь на одном примере. По-видимому, дифференциальные уравнения — это область математики, которая находит наиболее широкое применение у инженеров и ученых. Каждый, кто занимался этим предметом, помнит, насколько разобщенным он кажется; одно уравнение решается одним методом, другое уравнение — другим методом; бесчисленно количество типов уравнений, которые необходимо запомнить, бесчисленно количество методов решения. Как сказал один негодующий лектор: «Это же ботаника, а не математика!». (Замечание, довольно несправедливое по отношению к ботанике, которая, в конце концов, имеет целью классификацию, а не просто сбор образцов.) Сейчас имеется теория, разработанная Софусом Ли (1842—1893), которая устанавливает единый принцип, лежащий в основе всех типов дифференциальных уравнений [1]; эта теория показывает, почему они решаются именно используемыми методами. Несомненно, эта теория очень важна для любого математика, который хочет помочь практикам и показать, как решаются различные типы дифференциальных уравнений. Не случайно Софус Ли был геометром. Его идеи, оказавшиеся столь эффективными для дифференциальных уравнений, обязаны своим возникновением вопросам, тесно связанным с проективной геометрией.

ТЕОРИЯ ПЕРСПЕКТИВЫ

Мы сказали достаточно о красоте и полезности нашего предмета. Давайте же теперь проследим, как он развивается из теории рисунка.

---

[1] Это сказано, пожалуй, слишком сильно. Вообще, основные применения теории С. Ли (так называемой «теории групп Ли») связаны вовсе не с дифференциальными уравнениями.— *Прим. ред.*

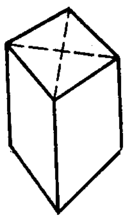


Рис. 50.

Вот рис. 50, который должен изображать куб. Рисунок во многом отличается от настоящего куба. Высота, ширина и длина куба равны между собой, между тем на рисунке это не так. У куба углы прямые, на нашем же рисунке соответствующие углы не прямые. Линии, которые идут параллельно друг другу в кубе, на нашем рисунке не параллельны. Не сохранено даже соотношение длин. Пунктирные линии, проведенные на верхней грани, в действительности делят друг друга пополам; на рисунке это не так.

Однако кое-что должно сохраниться. Если бы при переходе от объекта к его изображению ничего не сохранилось, то не существовало бы такого понятия, как хороший или плохой чертеж. И действительно, кое-что сохраняется. Например, если некоторая линия в действительности прямая, то на чертеже она тоже должна быть прямой. Изображение точки есть точка. Если точка лежит на прямой, то и на чертеже ее нужно показать лежащей на прямой. Так, глядя на фотоснимок, сделанный с самолета, вы могли бы сказать, что некоторая река протекает в действительности через такой-то город, потому что это вы видите на снимке. (Правда, города не точки, а реки не линии.)

Итак, фотография сохраняет следующие свойства: (1) быть прямой линией, (2) быть точкой, (3) «лежать на» (свойство точки и прямой). Эти свойства называются *проективными. Проективное свойство* — это такое свойство, которое сохраняется при фотографировании. Проективная геометрия занимается только такими свойствами, все остальные она игнорирует.

Если мы сфотографируем геометрическую фигуру, а потом сделаем фотографию с фотографии, то на второй фотографии все проективные свойства оригинала будут сохраняться. Не важно, сколько раз вы будете повторять этот процесс. Проективные свойства сохраняются на любой стадии процесса, следовательно, они сохраняются по окончании всего процесса.

Как мы уже видели раньше, такие вещи, как длина, угол, параллельность, отношение длин, изменяются при фотографировании. Все это не проективные свойства. Ни одно из них не рассматривается в собственно проективной геометрии.

Почти все результаты школьной геометрии касаются длин и углов. О чем же тогда говорить, если на эти вещи наложено табу? Существуют ли вообще тогда какие-либо теоремы?

На рис. 51 представлена проективная теорема, совершенно не зависящая от каких-либо измерений.

Из точки O проводим три прямые AO, OB, OC. Точки A, B, C могут быть расположены где угодно на этих прямых. Выделим еще любые три точки A', B', C': A' на OA, B' на OB, C' на OC. Соединим точки A и B, а также A' и B'; пусть F — точка пересечения пря-

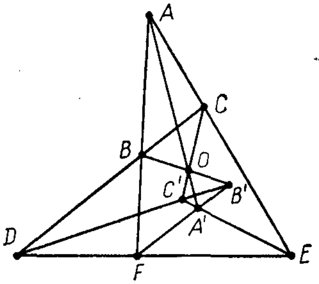


Рис. 51.

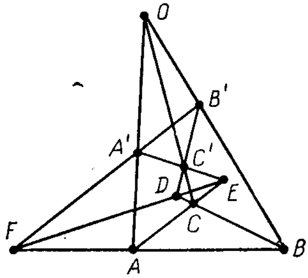


Рис. 52.

мых AB и $A'B'$. Аналогично AC и $A'C'$ пересекаются в точке E, а BC и $B'C'$ в точке D. Теперь вы можете обнаружить, что точки D, E, F лежат на одной прямой.

Этот результат не имеет никакого отношения к длинам или углам. Здесь используются только понятия прямой линии и точки, а также «лежать на». Теорема полностью проективна. Если вы сделаете фотографию этого чертежа, то на фотографии точки D, E, F тоже будут лежать на одной прямой.

Наш пример показывает, что существуют проективные теоремы.

Рассмотренный результат известен под названием *теоремы Дезарга* — того самого Дезарга, о котором упоминалось выше. Между прочим, эта теорема обладает замечательной симметрией. Каждая точка чертежа может занимать место любой другой. В построении, которое я привел выше, этот факт несколько скрыт. Мы начали с точки O, затем нанесли точки A, B, C и A', B', C' и, наконец, построили точки D, E, F. Но если вы теперь сотрете все буквы на чертеже, вы обнаружите, что можно любую из десяти точек обозначить буквой O. Тогда остальные девять точек вы сможете обозначить таким образом, чтобы сохранился смысл теоремы. Чертеж останется тем же самым, но точки будут обозначены другими буквами. На рис. 52 показан один из способов переименования точек. На самом деле существует 120 различных способов, которыми можно расставить буквы на этом чертеже, без какого-либо искажения смысла теоремы. Среди линий и точек чертежа царит полнейшая демократия. Имеется 10 линий, на каждой из которых расположено по три точки; имеется 10 точек, через каждую из которых проходят три линии. Такой чертеж называют «конфигурацией 10_3, 10_3».

Возникает совершенно естественный вопрос: какие еще чертежи обладают подобными свойствами? Иными словами: существуют ли и другие конфигурации? В конфигурации на каждой прямой должно располагаться одинаковое количество точек, а через каждую точку должно проходить одинаковое количество прямых.

Теорема Дезарга совершенно нечаянно дает ответ на детскую загадку: как посадить десять деревьев десятью рядами, так, чтобы в каждом ряду было по три дерева.

ПРОИСХОЖДЕНИЕ ТЕОРЕМЫ ДЕЗАРГА

Если вы хоть немного знаете евклидову геометрию, вы заметите, что эта теорема совершенно не похожа на большинство теорем школьной геометрии. Если бы от вас потребовали доказать ее методами евклидовой геометрии, то (хотя это и можно сделать) это была бы прескверная задача. Более интересным является вопрос о том, как Дезарг додумался до этого результата. После того как вы это узнаете, вы поймете, что результат очевиден.

Как уже говорилось, Дезарга интересовал вопрос, как делать чертежи зданий и других пространственных тел. Если вы посмотрите на теорему Дезарга как на чертеж пространственного тела, то результат вам сразу бросится в глаза.

Не всегда легко представить себе пространственное тело или отчетливо увидеть его по чертежу. Лучше всего сделать модель. Самый дешевый и подходящий материал для этого — старые газеты; каждый лист нужно свернуть в трубку толщиной с карандаш, но более длинную. Можно также воспользоваться сухими стеблями подсолнуха, если они у вас есть под рукой.

Возьмите три трубочки, сделайте из них треножник (рис. 53). Точка O расположена наверху, точки A, B, C — в основании.

Затем выберите точки A', B', C'. Как показано на рисунке, они размещены на трех ножках модели. Эти три точки не обязательно должны находиться на одном и том же расстоянии от основания. Лучше расположить их следующим образом: представьте, что какой-нибудь силач ударил острым мечом по треножнику через эти три точки, и меч, пройдя через модель, ударился о землю. Точки A', B', C' должны быть расположены так, чтобы меч ударился не очень далеко от точек A, B, C. Теперь положите одну трубочку так, чтобы она проходила через точки A' и B', а другую — на основание, так, чтобы она касалась точек A и B. Точкой пересечения этих трубок будет F. Аналогично, трубки AC и $A'C'$ пересекутся в точке E, а $B'C'$ и BC в точке D.

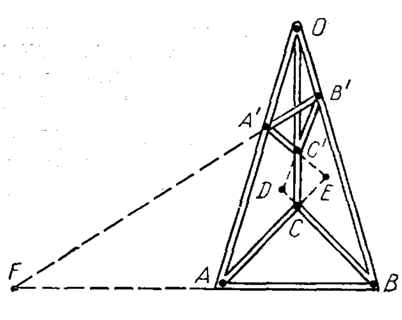


Рис. 53.

Найденные таким образом точки D, E, F все лежат на основании. Но все они также лежат в плоскости сечения мечом, о чем мы упоминали выше. Эти две плоскости пересекаются по прямой линии, как это видно из нашей модели. Если вы теперь сфотографируете эту модель или сделаете ее точный рисунок, вы получите чертеж теоремы Дезарга.

ТЕОРЕМА ПАППА

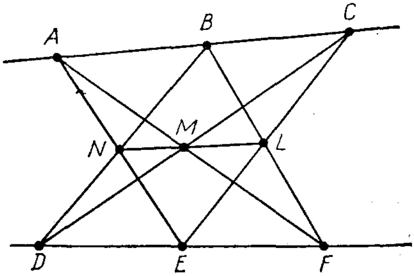


Рис. 54.

Еще один пример теоремы, не связанной с длинами и углами, представляет собой теорема Паппа. В ней говорится о девяти точках и девяти прямых, показанных на рис. 54. Естественно начать построение этого чертежа с того, чтобы провести две прямые — верхнюю и нижнюю — разместить точки A, B, C где-нибудь на верхней прямой, а точки D, E, F — на нижней. Там, где FB пересекает EC, ставим точку L, M — точка пересечения FA и DC. N — точка пересечения EA и DB. Оказывается, точки L, M, N лежат на одной прямой.

Этот чертеж также представляет собой конфигурацию, которую мы обозначаем 9_3, 9_3. Буквы на чертеже можно расставить 108 различными способами, причем смысл теоремы не изменится. Все точки равноправны.

ЧТО ТАКОЕ ИЗОБРАЖЕНИЕ?

Мы до сих пор говорили о «фотографировании» и об «изображении предметов», но мы не объясняли точно, что это значит. Представьте, что у вас в руках кусок стекла. Вы хотите на нем получить изображение различных предметов, расположенных позади него. Вы помещаете стекло на некотором постоянном расстоянии от глаза. Если вы хотите изобразить точку P предмета, вы делаете метку Q на стекле (рис. 55). Эта метка закрывает от вас точку оригинала; ваш глаз, точка Q и точка P лежат на одной прямой. Если вы нанесете таким образом на стекло метки для всех точек предмета, то вы получите его изображение.

На рис. 56 этот процесс показан в обратном направлении. Ли-

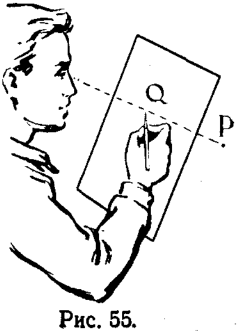


Рис. 55.

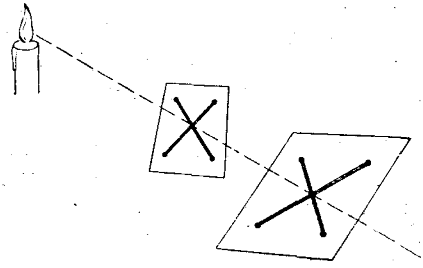


Рис. 56.

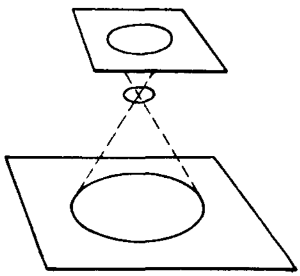


Рис. 57.

нии, изображенные на стекле, проектируются на экран, получается упрощенный вариант кинокамеры или волшебного фонаря.

Оба эти процесса мы будем называть *проектированием*. Мы уже привыкли употреблять в своей речи слово «кинопроектор», так что использование этого слова в связи со вторым рисунком нам кажется совершенно естественным. В геометрии неважно, уменьшается или увеличивается рисунок, поэтому мы будем использовать слово «проектирование» также и для первого рисунка.

Мы используем его и для ситуации, изображенной на рис. 57, где показано примерно то, что происходит в фотоувеличителе. Детали для нас не важны. Важно лишь то, что луч идет от точки к точке или через точку; рисунок получается там, где эти лучи пересекают плоскость.

ПРОЕКЦИЯ ПРЯМОЙ

Если на листе бумаги я отмечу любые две точки, я могу утверждать, что это снимок с самолета, показывающий положение Лондона и Парижа. Пусть L и P (рис. 58) указывают действительное положение Лондона и Парижа, а A и B — точки на моем листе бумаги. Мне нужно только расположить свой глаз в точке O, и тогда точка A скроет от моего взгляда Лондон, а точка B — Париж.

Итак, две точки на прямой вообще не обладают проективными свойствами; мы не получаем никакой информации, имея перед собой изображение двух точек, кроме того, что эти две точки различны.

Ничего также нельзя сказать о трех точках на прямой. Предположим, что вдоль прямой дороги расположены три города P, Q, R; я провожу линию с произвольно расположенными на ней точками A, B и C и утверждаю, что это аэрофотосъемка городов P, Q, R.

Представьте себе, что я устанавливаю свой рисунок так, что точка A совпадает с точкой P (рис. 59). Тогда, расположив глаз в точке O, я получу, что точки A, B и C будут служить изображениями P, Q и R. Как и раньше, я предполагаю, что P, Q и R — это различные точки, тогда A, B, C — тоже три различные точки.

Профессиональный математик, читая эти слова, отнесся бы к ним критически. Ведь на самом деле мы доказали только то, что три наперед заданные точки A, B и C служат изображениями P, Q и R при определенном «расположении рисунка» (т. е. прямой ABC). Но что здесь имеется в виду? Я должен логично объяснить, что подразумевается под «перемещением рисунка». (Чего, между прочим,

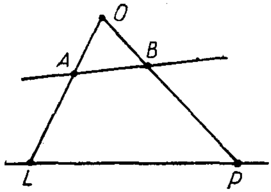


Рис. 58.

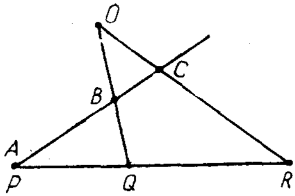


Рис. 59.

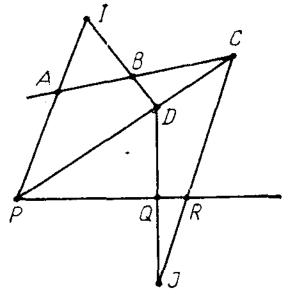


Рис. 60.

Евклид никогда не делал, и это одна из причин, по которой мы его теперь критикуем.) Однако я могу обойтись и без этого, сказав, что мой рисунок — это только фотография фотографии трех городов. Предположим, что A, B, C (рис. 60) — это предполагаемая фотография городов P, Q, R. Мы соединим точки P и C и ставим в любом месте полученной прямой точку D. Пусть PA пересекается с DB в точке I. Тогда, глядя из точки I, мы видим, что A, B, C есть правильное изображение P, D, C. Но P, D, C, в свою очередь,— это правильное изображение P, Q, R; чтобы убедиться в этом, нам нужно только посмотреть на вещи из точки J, где DQ пересекается с CR. Итак, ABC есть изображение PDC, а эта тройка точек, в свою очередь, является изображением PQR.

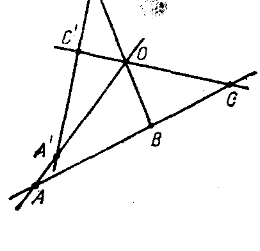
Но мы уже подчеркивали раньше, что изображение сохраняет все проективные свойства; эти свойства сохраняются и в изображении изображения. Так мы убеждаемся, что проективной геометрии нечего сказать о трех точках на прямой. Если бы где-нибудь на свете была громадная плоская пустыня и было бы известно, что там имеются три скалы, расположенные на одной прямой, то сделанная с самолета фотография, где были бы запечатлены эти три скалы и ничего больше, не содержала бы никакой новой информации [1].

Если положение самолета в момент фотографирования было бы неизвестным, то было бы невозможно что-либо сказать о расстояниях между скалами [2].

---

[1] Т. е. по этой фотографии можно было бы судить лишь о том, что скалы расположены на одной прямой; их взаимные расстояния оставались бы неизвестными.— *Прим. ред.*

[2] Возможно, что фотография позволила бы сказать, что скала B расположена *между* скалами A и C. Но это обусловливается природой фотокамеры. Если говорить о проектировании в самом общем смысле слова, имея в виду чистую геометрию, то при проектировании, вообще говоря, не сохраняется даже порядок трех точек на прямой. [Это можно видеть из прилагаемого рисунка, где точки A', B' и C' служат соответственно проекциями точек A, B и C из центра O.—*Ред.*]



Проективная геометрия ничего не может сказать о трех отдельных точках, расположенных на прямой, кроме того, что этих точек три и что они различные.

СЛОЖНОЕ ОТНОШЕНИЕ

Но положение меняется, когда на прямой задано четыре объекта. В этом случае аэрофотосъемка сообщает вам относительно них уже нечто определенное. Предположим, что A, B, C, D (рис. 61) — четыре точки на прямой. Примем в качестве начальной какую-нибудь точку O и предположим, что точки A, B, C, D расположены соответственно на расстояниях a, b, c, d справа от точки O.

Мы можем вычислить величину

$$x = \frac{(a-b)(c-d)}{(b-c)(d-a)}.$$

Это число обладает тем свойством, что *оно в точности одно и то же как для изображения, так и для оригинала.* Если вам нужно вычислить x, то не играет роли, измеряете ли вы расстояние на изображении или на самом участке.

Фотокамера может обмануть. Она обманывает, когда выдает равные длины за неравные и прямые углы за непрямые. Единственное, что она не искажает,— это выражение

$$\frac{(a-b)(c-d)}{(b-c)(d-a)}.$$

Значение этого выражения может быть найдено прямо из фотографии. И все, что можно с уверенностью утверждать, пользуясь свидетельством фотографии, может быть выражено в терминах таких величин.

Естественно, что эта величина имеет свое название. Оно называется *сложным отношением*[1] точек A, B, C, D. В качестве сокращенной записи этого выражения обычно используется символ $(ABCD)$[2].

Подчеркнем, что $(ABCD)$ означает одно число. Например, если на приведенном выше чертеже $a=2$, $b=3$, $c=5$, $d=7$, то $(ABCD)$ равно

$$\frac{(2-3)(5-7)}{(3-5)(7-2)} = \frac{(-1)(-2)}{(-2)\cdot 5} = -\frac{1}{5}.$$

Вы видите, что это число может быть и отрицательным. Но оно может быть и положительным, в чем нетрудно убедиться, положив, скажем, $a=2$, $b=3$, $c=5$, $d=4$.

---

[1] Или также *двойным отношением* четырех точек прямой.— *Прим. ред.*

[2] В конце главы 3, для сложного отношения чисел (координат точек), использовалось обозначение $f(a, b, c, d)$.— *Прим. ред.*

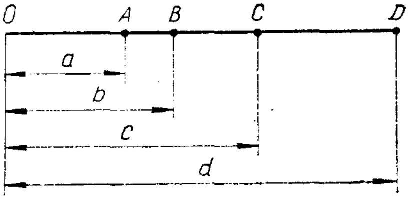


Рис. 61.

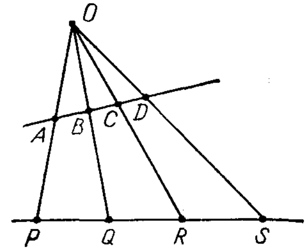


Рис. 62.

Замечательно, что выражение $(ABCD)$ было известно уже Паппу, который жил до 300 года н. э. Естественно, что он не мог выразить его ни в терминах фотографии (первое упоминание о которой датируется 1829 годом), ни в терминах перспективы (теория перспективы была развита лишь после 1400 года). Оно было связано с геометрической фигурой, изображенной на рис. 62, где $(PQRS) = (ABCD)$.

Неизвестно, открыл ли Папп этот результат сам или узнал его от какого-нибудь более раннего автора. Не знаем мы и хода мыслей, который привел к открытию этой теоремы. Об этом стоит очень и очень пожалеть, так как было бы чрезвычайно интересно знать, как она была открыта. Ибо это никоим образом не очевидный результат.

Представьте, что мы ничего не знаем об этом выражении и хотели бы найти такое свойство, которое было бы одинаковым как для изображения, так и для предмета. С чего бы мы начали поиски этого свойства? Современный математик, имея в своем распоряжении теорию инвариантов, знал бы, что делать. Но для человека, живущего в III веке,— это замечательное открытие.

Вот пример того, как сложное отношение может быть использовано в рисовании. На рис. 63 я попытался изобразить четыре фонарных столба, которые расположены вдоль дороги на равном расстоянии друг от друга. Правильно ли я нарисовал их?

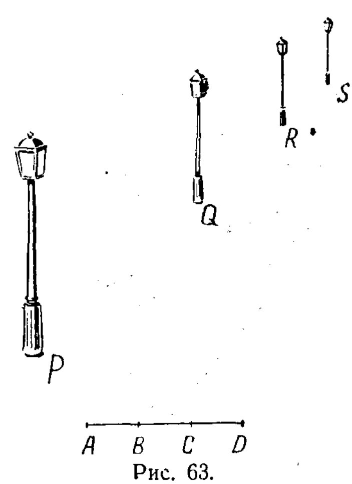


Рис. 63.

Если на этом рисунке я начну измерять расстояние от подножия ближайшего к нам фонаря, то я получу $p=0$, $q=6$, $r=8$, $s=9$. Следовательно, сложное отношение

($PQRS$) равно $\frac{(0-6)(8-9)}{(6-8)(9-0)} = -\frac{1}{3}$. Для четырех точек A, B, C, D, расположенных на равном расстоянии друг от друга, как на самом деле расположены фонарные столбы, мы имеем $a = 0$, $b = 1$, $c = 2$, $d = 3$; двойное отношение равно $\frac{(0-1)(2-3)}{(1-2)(3-0)}$, что снова дает $-\frac{1}{3}$. Если двойное отношение одинаково в обоих случаях, то это значит, что P, Q, R, S можно рассматривать как правильное изображение A, B, C, D.

Вы, конечно, понимаете, что эти вычисления вовсе не показывают, насколько правильно я нарисовал сами фонарные столбы, они лишь указывают, верно ли я расположил точки P, Q, R, S, являющиеся основаниями столбов.

ГОРИЗОНТ

Одна из первых вещей, которые узнает учащийся художественной школы,— это, как изображать параллельные прямые. На рис. 64 показано, какими кажутся нам на чертеже параллельные линии. Пусть наш чертеж изображает сквер с дорожками вокруг него. Фигура $ABCD$ изображает прямоугольник, каким он представляется глазу. Фигура $EFGH$ также представляет прямоугольник. AB, EF, GH, DC должны изображать параллельные прямые. Но учащийся не начертит их параллельными. Они будут начерчены таким образом, чтобы продолжения этих линий (показанные пунктиром) пересекались в точке Q. Аналогично предполагается, что AD, EH, FG, BC представляют собой параллельные прямые; на рисунке они пересекаются в точке P. Точки P и Q лежат на линии горизонта. Это особенные точки. Любая другая точка рисунка изображает реально существующую точку; E изображает угол сквера, D — поворот дорожки. Но точки P и Q не являются изображениями реальных точек. Нельзя, например, сказать: «У нас сегодня состоится пикник на том месте, которое представлено точкой Q!»

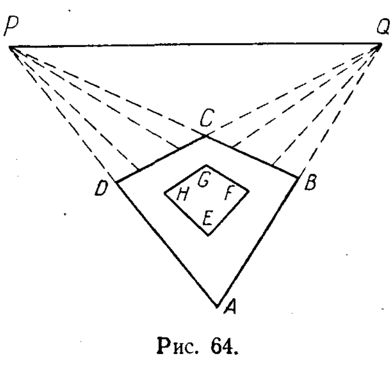


Рис. 64.

Самое большее, что мы можем сказать,— это то, что точка Q представляет *направление*. Любая прямая, начерченная на земле параллельно AB, должна пройти на рисунке через точку Q. Аналогично P представляет направление AD. Любая прямая на земле, параллельная AD, непременно пройдет на рисунке через точку P. Любая другая точка горизонта тоже представляет направление.

Итак, на чертеже мы имеем два рода точек: 1) настоящие точки, которые представляют реальные места, и 2) точки горизонта, в которых пересекаются (точнее, *кажутся* пересекающимися) параллельные прямые.

Линии, пересекающиеся в одной точке, носят название сходящихся. Так, на чертеже линии, которые на самом деле параллельны, выглядят сходящимися. Это означает, что параллельные и сходящиеся прямые должны иметь много общих свойств, и в действительности *все проективные свойства* будут общими для сходящихся прямых чертежа и для реальных параллельных прямых. В силу этого математики неизбежно должны были прийти к тому, чтобы рассматривать параллельные прямые как частный случай сходящихся («математика — это искусство давать одно и то же название различным вещам»). Мы привыкли говорить «параллельные прямые пересекаются в бесконечности»; можно считать, что P представляет «точку в бесконечности». Совокупность таких точек образует «прямую в бесконечности» [1], представленную на чертеже линией горизонта PQ.

Такой способ выражения дает нам возможность унифицировать теоремы. Вместо того чтобы сказать «две прямые либо пересекаются, либо параллельны», мы можем сказать «две прямые должны пересечься»; при этом, конечно, может случиться, что они пересекутся «в бесконечности»,— это новый способ выразить тот факт, что прямые не пересекаются.

В теореме Дезарга или теореме Паппа прямые, которые я раньше считал пересекающимися, могут оказаться параллельными. Но тогда я буду считать, что они пересекаются в бесконечности, и все равно получу верный результат. Сделайте чертежи к этим теоремам так, чтобы некоторые прямые были параллельны, и вы увидите, что случится с теоремами.

Возможно, вы сомневаетесь, имеет ли смысл говорить о точках в бесконечности как о чем-то действительно существующем. Это очень хорошо, если у вас возникают такие сомнения, они способствуют формированию мировоззрения. Позже я буду говорить о логике предмета, и вы посмотрите, удовлетворит ли вас моя точка зрения. Пока мы просто исследуем. Мы лишь пытаемся говорить о бесконечности как о реально существующей вещи и смотрим, куда нас это приводит. Если результаты нам понравятся, мы будем считать, что так можно говорить.

ОТНОШЕНИЕ ДЛИН В ИЗОБРАЖЕНИИ

Представьте, что мы смотрим на чертеж, где изображены три точки A, B, C, расположенные на одной прямой; они представляют три камня, лежащие на земле. Можем ли мы, глядя на этот чер-

[1] Более распространенные термины: «бесконечно удаленная точка», «бесконечно удаленная прямая».—*Прим. ред.*

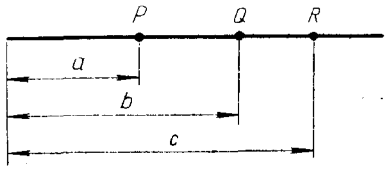


Рис. 65.

теж, заключить что-либо о действительном расположении камней?

Ясно, что, имея только три точки A, B, C, мы ничего не сможем заключить, потому что, как мы уже видели раньше, в проективной геометрии три точки на прямой ничем не отличаются от любых других трех точек; они не имеют проективных свойств. Но предположим, что на чертеже указан горизонт и наша прямая AC пересекает его в точке D, так что D представляет точку в бесконечности для линий камней. Теперь у нас имеются уже четыре точки A, B, C, D на прямой, и они нам кое о чем расскажут.

Двойное отношение $(ABCD)$ на чертеже равно двойному отношению точек на земле. Пусть a, b, c будут расстояния до точек A, B, C от некоторой произвольно фиксированной точки на этой же прямой. Поскольку D представляет точку в бесконечности на этой прямой, то мы должны принять $d = \infty$. Ни один уважающий себя аналитик и не подумает заменить символ бесконечной величиной, и тем более он не сделал бы то, что мы сейчас собираемся сделать. Однако *в проективной геометрии* такие вещи оправданы[1]. Мы подставляем $d = \infty$ в выражение $\frac{(a-b)(c-d)}{(b-c)(d-a)}$; тогда $d - a = \infty$, в то время как $c - d = -\infty$. Сокращая дробь, находим, что двойное отношение равно $-\frac{a-b}{b-c}$. Но это все равно, что $-\frac{(b-a)}{(c-b)}$. Если P, Q, R (рис. 65) представляют действительное положение камней на земле, то $(b-a)$ — это расстояние PQ, а $(c-b)$ — расстояние QR, следовательно, $-\frac{(b-a)}{(c-b)} = \frac{PQ}{QR}$.

Таким образом, двойное отношение точек P, Q, R и бесконечно удаленной точки на этой прямой равно $-\frac{PQ}{QR}$; иными словами, это есть отношение, в котором Q делит отрезок PQ, притом взятое с отрицательным знаком. Но двойное отношение в точности сохраняется при переходе от оригинала к чертежу. Следовательно, если на чертеже мы измерим расстояния AB, AC, AD и подсчитаем двойное отношение $(ABCD)$, то оно будет иметь знак минус, а его абсолютная величина скажет нам, в каком отношении средний камень делит отрезок, соединяющий два других камня.

В частности, если Q — середина отрезка PR, то $PQ = QR$,

[1] В проективной геометрии есть бесконечно удаленная прямая, в комформной геометрии — бесконечно удаленная точка; в анализе же бесконечность есть табу. Таким образом, слово «бесконечность» в этих трех случаях имеет разный смысл, и очень жаль, что у нас есть только одно слово для выражения столь различных понятий.

$\frac{PQ}{QR} = 1$ и $(ABCD) = -1$. Говорят, что четыре точки A, B, C, D на прямой, для которых $(ABCD) = -1$, образуют *гармонический ряд*[1].

Итак, грамонические ряды естественным образом возникают в связи с теорией перспективы. Гармонический ряд служит изображением разделенного пополам отрезка прямой вместе с бесконечно удаленной точкой прямой. Гармонические ряды играют значительную роль в проективной геометрии, понимаемой в современном смысле этого слова (как предмет, не зависящий ни от теории перспективы, ни от понятия длины).

ПОСТРОЕНИЕ ГАРМОНИЧЕСКИХ РЯДОВ

Выше мы рассказали о том, как сделать чертеж сквера. Предположим, что наш сквер $ABCD$ (рис. 66) имеет форму прямоугольника. AC и BD пересекаются в точке R, RS параллельна AD. Ясно, что S есть середина AB.

Теперь предположим, что мы сделали чертеж сквера. Он будет выглядеть следующим образом (рис. 66). Точки P и Q лежат, разумеется, на линии горизонта. Поскольку $ASBQ$ есть изображение разделенного пополам отрезка и точки в бесконечности, $ASBQ$ должно быть гармоническим рядом, или $(ASBQ) = -1$.

Теперь мы можем забыть о сквере и смотреть на последний чертеж как на геометрическую фигуру, а не как на изображение чего бы то ни было. Он дает нам способ построения гармонического ряда. Чтобы нарисовать этот чертеж, нужна только линейка, чертеж состоит из одних прямых линий; нам не приходится измерять длины или углы. Это чисто проективное построение гармонического ряда[2]. Если мы сделаем фотографию такого чертежа, то она будет так же нас устраивать, как сам чертеж.

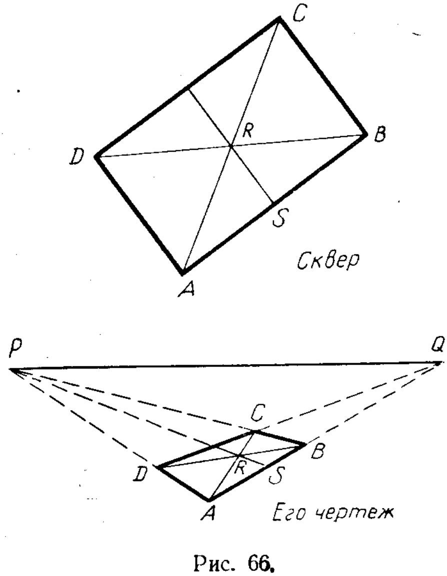


Рис. 66.

[1] [Или *гармоническую четверку точек*.— *Ред.*] Название «гармонический» обусловлено тем, что длины AB, AC, AD играют некоторую роль в теории музыкальных инструментов.

[2] Для большей отчетливости укажем еще раз это построение. Точки A, B, C, D берутся произвольно. Затем находятся точки P и R и, наконец, точки S и Q. Четверка точек A, S, B, Q образует гармонический ряд.— *Прим. ред.*

ПРИМЕНЕНИЕ ГЕОМЕТРИИ К АЛГЕБРЕ

В главе третьей упоминалось сложное отношение, но там предполагалось, что читателю нужно проделать ряд довольно громоздких алгебраических преобразований, чтобы проверить его свойства. Сейчас мы возвращаемся к тому же вопросу вооруженные новыми знаниями. Мы знаем, что сложное отношение есть нечто такое, что остается неизменным при проектировании. Мы знаем и то, что три точки могут быть спроектированы в любые три точки. Это значительно упростит работу.

Мы будем проектировать три точки в стандартные положения, выбранные с таким расчетом, чтобы сделать выражения по возможности более простыми. Как же наилучшим образом выбрать эти три точки, или три числа, фиксирующие положение точек? Очевидно, наш выбор должен пасть на число 0, а также на число 1. С этими числами очень легко производить вычисления. Третье выбранное нами число — бесконечность, или ∞. В эти положения мы будем проектировать наши три точки.

Пусть A, B, C, D — четыре точки на прямой (рис. 67), расположенные на расстояниях a, b, c, d от некоторой фиксированной точки на этой же прямой.

Через точку A проводим прямую AX. Пусть точка N выбрана на AX таким образом, что расстояние AN равно 1. Соединим N и D. Через точку C проведем прямую, параллельную AX, пересекающую ND в точке O. O будет точка, из которой мы проектируем, так, чтобы тени от точек A, B, C, D падали на прямую AX. Поскольку OC параллельна AX, C проектируется в бесконечность, как мы этого и хотели. Точка A уже лежит на AX и проектируется сама в себя. D проектируется в N. Если мы будем измерять расстояния от точки A, то сама точка A будет расположена на расстоянии O, N — на расстоянии 1. Обозначим буквой Q проекцию точки C; Q будет бесконечно удаленная точка прямой AX, ее расстояние от A равно ∞. И, наконец, проекция точки B есть M. Обозначим расстояние AM через X.

Поскольку сложное отношение не изменяется при проектировании, то

$$(ABCD) = (AMQN),$$

или, в числовых обозначениях, $f(a, b, c, d) = f(0, x, \infty, 1)$. Но $f(0, x, \infty, 1)$ легко подсчитать. Обращаясь с символом ∞ как раньше, мы получим

$$f(0, x, \infty, 1) = \frac{(-x)(\infty)}{(-\infty)(1)} = x.$$

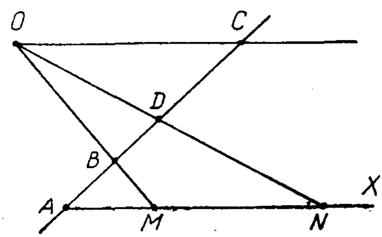


Рис. 67.

Аналогичным образом, изменив порядок букв, получим, например,

$$(ABDC) = (AMNQ),$$

или

$$f(a,\ b,\ d,\ c) = f(0,\ x,\ 1,\ \infty) = \frac{(-x)(-\infty)}{(x-1)(\infty)} = \frac{x}{x-1}.$$

Это один из результатов, с которым мы встретились в главе третьей, но на этот раз мы получаем его очень легко, геометрически, без каких-либо громоздких алгебраических преобразований, и нам не нужно думать, какой способ является наилучшим для проверки результата.

Пример убедительно показывает нам, как удобно знать, что некоторая функция остается неизменной при выполнении определенной операции; в данном случае такой операцией является проектирование.

Поэтому при изучении любой задачи разумно бывает спросить: «Какие операции оставляют ее без изменения?». И чем больше мы можем найти подобных операций, тем большей свободой будем мы обладать в приведении данной задачи к удобному виду. В последнем примере свобода состояла в том, что мы могли придать трем из четырех символов, встречающихся в выражении $f\ (a,\ b,\ c,\ d)$, три заранее выбранных нами значения, а именно, 0,1 и ∞.

Вывод, к которому мы пришли, имеет очень широкое применение.

Функция, остающаяся неизменной при выполнении некоторой операции, называется *инвариантной* по отношению к этой операции. Так, $f(a,\ b,\ c,\ d)$ *инвариантна относительно проективного преобразования.*

Но проектирование — это все-таки геометрическое понятие. Если нам удастся перевести его на язык алгебры, мы получим чисто алгебраический результат, нечто такое, что мы можем проверять алгебраически без всякого обращения к геометрическим построениям.

На рис. 68 приведен пример проектирования. Прямые OX и HJ, а также точка E фиксированы; точка F движется вдоль прямой HJ. Если считать, что в точке E помещен источник света, то тень, отбрасываемая точкой F на прямую OX, дает точку G.

Если мы обозначим расстояние HF буквой t, а расстояние OG — буквой u, то величина u получается из t с помощью геометрического построения, включающего проектирование, т. е. с помощью проективного преобразования. Сейчас наша цель состоит в том, чтобы найти алгебраическую зависимость u от t.

Но отыскание u как функции t — это несложная задача из аналитической геометрии. Мы просто должны проследить поэтапно геометрическое построение и на каждом этапе переводить происходящее на язык алгебры.

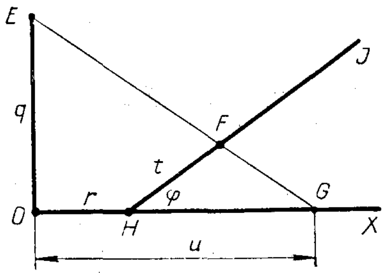


Рис. 68.

Мы предполагаем, что O — начало координат, E — точка с координатами $(0, q)$, H — точка с координатами $(r, 0)$, а прямая HJ образует угол φ с осью OX. Все сказанное обозначено на чертеже. Величины q, r, φ, конечно, постоянные. Меняются только t и u. Если хотите, вы можете считать, что t обозначает время, так что точка F движется вдоль HJ согласно закону $s=t$, т. е. с постоянной скоростью. Точка G, будучи тенью F, тоже движется, и мы хотим найти закон ее движения.

В любой момент времени t точка F имеет координаты

$$(r+t\cos\varphi,\ t\sin\varphi),$$

что можно легко увидеть, опустив перпендикуляр из точки F на HG. Теперь нам известны координаты точек E и F. Но через две точки проходит только одна прямая, и нам остается проделать шаблонные преобразования, чтобы найти ее уравнение. Это уравнение можно найти несколькими различными способами; имеются разные методы и формулы. Поскольку мы уже рассматривали этот вопрос в главе четвертой, мы можем с полным правом использовать уравнение (3) этой главы, куда мы должны подставить $a=r+t\cos\varphi$, $p=t\sin\varphi$ в качестве координат точки F и $b=0$ в качестве первой координаты точки E. Буква q уже имеет нужное значение, так она является второй координатой точки E. Итак, уравнение прямой EF имеет вид:

$$y = x\,\frac{t\sin\varphi - q}{r + t\cos\varphi} + q.$$

Точка G лежит на пересечении этой прямой с осью OX, уравнение которой есть $y=0$. Подставляя $y=0$ в полученное уравнение и разрешая его относительно x, мы находим, что первая координата точки G должна быть равна

$$u = \frac{q\,(r + t\cos\varphi)}{q - t\sin\varphi}.$$

Этот результат выглядит гораздо более сложным, чем он есть на самом деле. Единственная переменная здесь t, и нас интересует лишь роль буквы t в этом уравнении (это же самое нас интересовало в главе четвертой). Если мы выберем конкретные численные значения констант q, r, φ, то простота формулы станет очевидной. Мы можем, например, выбрать значение угла φ равным $53°8'$[1];

[1] Конечно, это *приближенное* значение $\arcsin\frac{4}{5}$. — *Прим. ред.*

косинус и синус этого угла очень просты: $\cos\varphi = \frac{3}{5}$, $\sin\varphi = \frac{4}{5}$. Принимая затем $q = 1$, $r = 2$, мы получим пример проективного преобразования:

$$u = \frac{10 + 3t}{5 - 4t}.$$

При других значениях q, r, φ числа не были бы арифметически столь же простыми, но они все равно оставались бы константами, и мы всегда имели бы соотношение вида

$$u = \frac{g + ht}{R + lt} \quad \text{(ПП)}.$$

Приведенному выше примеру соответствуют значения $g=10$, $h=3$, $k=5$, $l=-4$. Буквы ПП означают «Проективное преобразование». Мы видим, что геометрическому понятию «проектирование» соответствует очень простой алгебраический процесс.

Таким образом, геометрия обращает наше внимание на тип алгебраической функции, обозначенный буквами ПП. Эта функция обладает важными свойствами и встречается во многих областях математики, которые не имеют явной связи с геометрией.

Например, представьте, что мы изучаем интеграл

$$\int \frac{dx}{\sqrt{(x-a)(x-b)(x-c)(x-d)}}.$$

Очень часто интеграл можно взять, полагая $x=F(z)$; если умело выбрать $F(z)$, то интеграл можно значительно упростить. Это хорошо известный из курса математического анализа способ замены переменной. Но какой должна быть функция $F(z)$? Здесь часто возникает затруднение.

Если мы посмотрим на интеграл, то увидим, что в него входят четыре числа a, b, c, d. Очевидно, что-то особенное происходит, когда $x=a$, $x=b$, $x=c$, $x=d$. Мы уже видели раньше, что любые три точки можно спроектировать в любые другие три точки. Алгебраически этот результат звучит так: всегда можно найти такое проективное преобразование, которое переводит любые три числа в любые другие три числа. Обычно выбираются, как мы видели, число 0,1 и ∞. Это означает, что мы можем выбрать константы g, h, k, l в уравнении

$$x = \frac{g + hz}{k + lz}$$

таким образом, чтобы было $x=a$ при $z=0$, затем $x=b$, когда z равно (или стремится к) бесконечности, и $x=c$ при $z=1$.

Отыскание констант не представляет трудности, но я не буду проделывать этого здесь. Мы предположим, что константы найдены и подставлены в верхнее уравнение. Затем мы используем это уравнение для замены переменной x в интеграле на переменную z.

Выполнение этой работы требует только знаний из элементарной алгебры и дифференциального исчисления, а также готовности провести довольно громоздкие преобразования. Никаких принципиальных затруднений в этой работе не возникает.

Для того, кто не желает заниматься вычислениями, представляет интерес сама идея; решения некоторых задач математического анализа могут быть значительно упрощены за счет использования алгебраической функции, взятой из геометрической теории аэрофотосъемки. Это блестящим образом иллюстрирует взаимозависимость различных разделов математики и еще раз подчеркивает тот факт, что существует предмет *математика,* а не просто набор методов, имеющих инженерное применение.

После кропотливых вычислений рассматриваемый интеграл запишется в следующем упрощенном виде:

$$C\int\frac{dz}{\sqrt{z(z-1)(z-f)}}.$$

Под знаком корня теперь стоит выражение третьей степени вместо четвертой. (Это обусловлено тем, что одна точка отправилась в бесконечность.) Величина f, входящая в подынтегральную функцию, есть как раз то, что мы раньше обозначали $f(a, b, c, d)$, т. е. сложное отношение чисел a, b, c, d. И снова геометрия помогает нам: с выражением $f(a, b, c, d)$ мы встречались уже в геометрии. Поэтому мы узнаем эту функцию, когда она появляется в задаче математического анализа. Если бы мы раньше не занимались проективной геометрией и, следовательно, не понимали бы значения сложного отношения, мы бы не обратили внимания на выражение $\frac{(a-d)(b-c)}{(a-b)(c-a)}$, когда оно нам встречается в процессе вычислений. Оно казалось бы нам еще одним собранием символов. А сейчас мы говорим: «А, сложное отношение! Откуда оно здесь взялось?».

ПРЕОБРАЗОВАНИЯ

Рассмотренный пример показывает, насколько полезны преобразования для классификации задач. Два интеграла из последнего раздела выглядят совершенно различными. В одном из них под знаком корня стоит многочлен четвертой степени, в другом — третьей. Но с помощью проективного преобразования нам удалось привести первый интеграл ко второму; при желании мы могли бы проделать это в обратном порядке, т. е. перейти от второго к первому. Поэтому каждый интеграл можно рассматривать как *скрытую форму другого*. Все, что мы знаем об одном из них, в равной степени относится и ко второму. Мы начинаем рассматривать их не как две различные задачи, а как две формы одной и той же задачи.

В этом и состоит великое значение преобразований. Задачи, которые с первого взгляда кажутся различными, превращаются в одинаковые. Решив задачу, мы не просто узнаем решение именно этой задачи, но также решения *всех задач, которые можно рассматривать как замаскированную форму той же самой задачи.* Иногда бывает, что преобразования не столь всемогущи. Часто они выявляют лишь *одну* разновидность для каждой задачи. В этом случае они удваивают наши знания. К каждой уже решенной задаче они подбирают пару. Таким образом, задачи классифицируются по парам. Но бывают и более мощные преобразования. Они могут содержать одну или более констант, которые мы можем выбирать по желанию. Тогда мы можем привести нашу задачу к бесчисленному множеству различных форм. Узнавая ее решение, мы узнаем решение всего семейства задач.

Но и здесь математика не останавливается. Мы нашли одно преобразование — проективное, которое представляет определенную ценность и интерес. Вначале, возможно, нам покажется достаточным использование этого типа преобразования, позволившего нам классифицировать задачи. Но позже нам приходит в голову мысль: «Этот частный тип преобразования помог нам. Он умножил наши знания. А какие еще типы преобразований существуют?».

Мы хотим не просто обобщить наши *результаты;* мы хотим также обобщить и наши *методы.* Найдя некоторый метод полезным, мы ищем другие методы, которые обладали бы подобными свойствами. Найдя полезным какое-нибудь преобразование, мы ищем другие преобразования, мы спрашиваем, каковы их свойства, мы пытаемся их классифицировать. Очевидно, что изучение преобразований — это ценное направление исследований.

Математический процесс бесконечен. Сначала мы классифицируем сами задачи; это достаточно легко понять. Затем мы классифицируем способы классификации задач; это звучит несколько сложнее. Дальше следует третий этап процесса, но это уже слишком похоже на скороговорку, чтобы стоило продолжать. И, наконец, четвертый и пятый этапы, каждый из которых классифицирует то, что уже было достигнуто раньше.

Этот процесс не имеет конца. Но каждая стадия процесса приносит определенное удовлетворение: она обогащает нас рядом принципов и помогает охватить более широкий круг знаний. К изучению преобразований мы вернемся в главе 12.

Эту главу мы начали с упоминания об окружностях, имеющих две общие точки. Но мы не обсудили этот вопрос, равно как и вопрос о логической оправданности существования прямой в бесконечности. Этими вопросами мы займемся в следующей главе.

Глава одиннадцатая

О КАЖУЩИХСЯ НЕВОЗМОЖНОСТЯХ

В душе у каждого поэта
Живет своя луна.
Но та, чей образ до рассвета
Хранит, зыбясь, волна,
Та, что таит свои секреты
От астрономов,— но приветы
Влюбленным шлет,— одна!

Ральф Ходжсон[1]

Стоит только показать, что какая-либо вещь невозможна, как найдется математик, который ее сделает.

Из разговоров

В предыдущей главе мы пользовались таким довольно сомнительным понятием, как бесконечность, и даже позволяли себе разного рода философские вопросы, вроде «*Действительно* ли существует место, именуемое бесконечностью, где пересекаются железнодорожные рельсы? И если такового нет, то имеют ли математики право говорить о бесконечности, где пересекаются параллельные линии?»

Ответом на оба вопроса может быть и (1) «Нет», и (2) «Да». В том, что на второй вопрос следует ответить утвердительно, я уверен: употребление в математике понятия «прямая в бесконечности» оправдано. Первый вопрос относится скорее к физике, чем к математике, это вопрос о том, чтó происходит на границе вселенной. Я не знаю, существуют ли вообще границы вселенной; если да, то я там никогда не был и потому не знаю, что они собой представляют, и как вели бы себя железнодорожные рельсы, если бы их перенесли туда.

Возможно, этот ответ оставляет у вас такое ощущение, что вас в чем-то обманывают. Мы привыкли думать, что геометрия — это наука, которая говорит нам правду. Как же можно говорить о геометрии, как о правдивой науке, если многое в природе происходит совсем по-другому? Такая постановка вопроса может смутить даже математика. Мы отнюдь не хотим, чтобы вы подумали, что математика не имеет *никакого* отношения к истине.

В математике на первом плане стоит логичность доказательств. Математик может сказать физику: «Вот последовательная теория. Я не знаю, применима ли она к реальной физической вселенной, но я уверен, что она последовательна сама по себе; она ответит «да» или «нет» на любой предложенный ей вопрос и никогда не ответит и «да», и «нет» на один и тот же вопрос».

---

[1] Перевод Ю. А. Гастева.— *Прим. перев.*

Задачей физика является проверить, согласуется ли подобная математическая теория с наблюдаемыми фактами; совпадают ли математические закономерности с закономерностями, встречающимися в реальной жизни.

Оказывается, что такой подход чрезвычайно плодотворен и что математические теории в действительности приводят к практическим результатам. Я, правда, не уверен, что все может быть предсказано только математикой. Вселенная могла бы быть и хаотичной. Тот факт, что она упорядочена, что ее можно постигнуть средствами логического мышления, является, по-моему, выводом из данных опыта и не поддается логическому или математическому доказательству.

Когда мы говорим, что математическая теория оправдана, мы имеем в виду только то, что эта теория, сколько бы вы ее ни применяли, не приведет вас к противоречиям. Мы утверждаем только, что она *формально непротиворечива,* и ничего больше. Но даже это, как мне кажется, является утверждением реальности теории; ведь мы говорим, что, сколько бы мы ни пользовались этой теорией, мы никогда не будем противоречить сами себе (при условии, конечно, что мы применяем ее правильно). Конечно, если в этом утверждении со временем будет обнаружена ошибка, нам придется отказаться от уверенности в правильности наших методов.

В этой главе я попытаюсь показать, что можно построить непротиворечивую систему, в которой параллельные прямые пересекаются в бесконечности. Непротиворечивость этой системы признают все живущие ныне математики[1]. Вам судить, найдется ли в будущем в этой системе какой-нибудь изъян.

ОДНОРОДНЫЕ КООРДИНАТЫ

Мы вскоре увидим, что о точках в бесконечности гораздо легче говорить, используя особую конструкцию, которая носит название однородных координат. Эта конструкция оказывается очень полезной как в проективной геометрии, так и в других областях математики. Возможно, на первый взгляд эта конструкция покажется вам не особенно естественной, вы даже подумаете, что она излишне усложнена; ибо вместо того, чтобы обозначать точку на плоскости,

---

[1] Во всяком случае, практически все. Для некоторых математиков-философов все еще является спорным вопросом непротиворечивость арифметики; если же признать, что арифметика непротиворечива, то отсюда автоматически следует непротиворечивость проективной геометрии. [Интуитивная *убежденность* «практически всех» математиков в непротиворечивости арифметики (и даже анализа!) отнюдь не снимает *проблему непротиворечивости,* которая формулируется в *точных,* а не каких-либо «философских» терминах — как и *любая* математическая проблема. Уместнее поэтому было бы не противопоставлять эти два подхода, а просто говорить, как это обычно и делается, об *относительной непротиворечивости,* т. е. непротиворечивости геометрии (и не только проективной!) *относительно* арифметики.— *Ред.*]

как обычно, двумя числами (x, y), мы используем три числа (X, Y, Z). Однако в плоской геометрии во многих случаях для симметрии алгебраических выражений желательно иметь три координаты вместо двух. (К сожалению, примеры требуют некоторого углубления в детали, а это здесь нежелательно.)

Новые числа (X, Y, Z) связаны со старыми координатами (x, y) достаточно простыми уравнениями:

$$x = X/Z, \quad y = Y/Z.$$

Например, если мы хотим записать координаты точки $x = \frac{7}{8}$, $y = \frac{3}{8}$ в новой системе, мы можем принять $X = 7$, $Y = 3$, $Z = 8$. Подставив эти значения в правые части приведенных выше уравнений, мы получим желаемые значения x и y. Я говорю «мы можем», а не «мы должны» выбрать значения 7, 3, 8 для X, Y, Z: дело в том, что $\frac{7}{8}$ все равно, что $\frac{14}{16}$, и $\frac{3}{8}$ все равно, что $\frac{6}{16}$, поэтому мы могли бы с таким же успехом выбрать числа 14, 6, 16 или любые другие три числа вида $7k$, $3k$, $8k$.

Это первая особенность однородных координат: вы можете умножить величины X, Y, Z на любое число, не меняя при этом положения точки, которую они представляют.

Теперь мы можем уже продемонстрировать некоторые упрощения, к которым приводят новые координаты. На рис. 69 точка A имеет координаты $x=2$, $y=1$, а точка B — $x=8$, $y=4$. В новой системе мы могли бы принять для точки A координаты (2, 1, 1), а для точки B (8, 4, 1). Предположим, нам теперь вздумалось сложить эти числа; получим ли мы что-нибудь интересное? Под сложением я понимаю здесь следующее: числа, определяющие точку B, пишутся под числами, определяющими точку A, после чего производится сложение; в результате получаются три отдельные суммы:

$$\begin{array}{lr} & A\ (2,\ 1,\ 1), \\ & \underline{B\ (8,\ 4,\ 1),} \\ \text{После сложения получаем} & (10,\ 5,\ 2). \end{array}$$

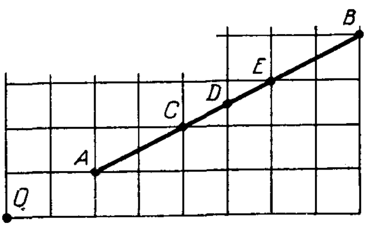


Рис. 69.

Это действие выполнить нетрудно. Но имеет ли оно какой-либо геометрический смысл? Если мы вернемся к обычной системе координат, то мы увидим, что числа 10, 5, 2 соответствуют точке с координатами $x = 5$, $y = 2\frac{1}{2}$. Это координаты точки D — — середины отрезка AB.

Я всегда представляю себе точку D как своего рода «смесь» точек A и B — вроде настоя, изготовленного из чая двух сортов, взятых в равных количествах. Посмотрим, что произойдет, если мы изменим пропорцию. Положим, мы берем два раза A и один раз B.

Удвоенные	координаты точки A	(4, 2, 2),
	координаты точки B	(8, 4, 1),
При сложении получаем		(12, 6, 3).

Этот результат соответствует точке C с координатами $x=4$, $y=2$, лежащей на расстоянии $\frac{1}{3}$ пути от A до B. Новая точка по-прежнему лежит на прямой AB, но удвоение вклада со стороны точки A перетянуло новую точку ближе к A. Мы можем без труда убедиться в том, что, оставив без изменений координаты (2, 1, 1) точки A, но удвоив координаты точки B, мы получим после сложения координаты новой точки E, которая будет расположена на расстоянии, равном двум третям пути от A до B.

Естественным и кратчайшим способом записи этих результатов является следующий: $D=A+B$, $C=2A+B$, $E=A+2B$. Как вы можете проверить сами, $A+\frac{1}{2}B$ тоже дает точку C. Настой чая, полученный из смеси 200 *г* цейлонского и 100 *г* китайского чая (если подобные смеси вообще кто-либо захочет делать), будет таким же, как настой из 100 *г* цейлонского и 50 *г* китайского чая[1].

Однако применять эти обозначения нужно с осторожностью. Не делайте, например, следующей ошибки: $2A+OB$ приводит к точке A, значит, $2A=A$; прибавляя B к обеим частям этого равенства, получим $2A+B=A+B$, т. е. $C=D$. Но C и D — это две различные точки! Ошибочность такого рассуждения ясно видна на примере с чаем. Чашка чая, приготовленная из 200-граммовой пачки цейлонского чая, имеет тот же вкус, что и чашка чая, заваренного из 100-граммовой пачки. Но отсюда не следует, что чай, заваренный из смеси 200 *г* цейлонского и 100 *г* китайского чая, будет иметь одинаковый вкус с заваркой 100 *г* цейлонского и 100 *г* китайского чая.

Имея дело с такими уравнениями, нужно использовать только такие рассуждения, которые верны для пропорций смесей. Пропорции 1 : 0 и 2 : 0 образуют смесь одного типа (в ней содержится только первый ингредиент), но пропорции 1 : 1 и 2 : 1 дают совершенно различные смеси.

Возвращаясь к нашему рисунку, мы видим, что точки $A+B$, $2A+B$, $A+2B$ лежат на отрезке AB. Выбрав соответствующим образом положительные числа m и n, мы можем получить точку $mA+nB$ в любом месте отрезка AB. (Возможно, вы сочтете полезным проделать самостоятельно несколько примеров. Попробуйте,

---

[1] Разумеется, если в первом случае взять вдвое большее количество воды.— *Прим. ред.*

например, найти, где находится точка $99A+B$ или $29A+51B$.) Чтобы выйти за пределы отрезка AB, нам придется пользоваться отрицательными числами. Например, если мы возьмем $-4A+B$, то получим следующее:

$$\begin{array}{rlr} -4 \text{ раза} & \text{координаты точки } A & (-8,\ -4,\ -4), \\ & \text{координаты точки } B & (\ \ 8,\ \ \ 4,\ \ \ 1), \\ \hline & \text{результат сложения} & (\ \ 0,\ \ \ 0,\ -3). \end{array}$$

Итак, $x=0$, $y=0$. Выражение $-4A+B$, таким образом, представляет начало координат — точку O, которая, как вы видите из чертежа, тоже лежит на прямой AB.

Если вы определите координаты точек $-3A+B$, $-2A+B$, $-1\frac{1}{2}A+B$, $-1\frac{1}{4}A+B$, $-1\frac{1}{8}A+B$, то увидите, что все эти точки лежат на продолжении отрезка AB, проходящем через O. Каждая последующая точка лежит левее предыдущей. Коэффициент при A каждый раз ближе к -1. Если мы возьмем его равным -1 и рассмотрим точку $-A+B$, то получим в качестве координат (X, Y, Z) числа $(6, 3, 0)$. В этом случае мы не в состоянии найти x, y, так как наши стандартные уравнения $x=\frac{X}{Z}$, $y=\frac{Y}{Z}$ приводят к бессмысленным выражениям $6/0$, $3/0$. Как только мы проходим значение -1, все затруднения исчезают, мы опять можем без всякого труда найти значения x и y. Рассматривая последовательность величин -4, -3, -2, $-1\frac{1}{2}$ и т. д., приближающихся к -1, мы замечаем, что соответствующие им точки смещаются вдоль прямой все дальше и дальше влево. Подойдя достаточно близко к -1, мы можем получить точку, расположенную сколь угодно далеко от 0. А как вы думаете, что случится, когда значение -1 будет пройдено? Где лежит точка $-\frac{7}{8}A+B$ и где находится точка $-\frac{1}{2}A+B$? (Поскольку для ответов на эти вопросы нужна только арифметика, я предоставляю вам все проделать самим. Очень интересно проследить путь точки $kA+B$, где k меняется от большого отрицательного числа, скажем от $-1\,000\,000$, до большого положительного значения, например $+1\,000\,000$, проходя через значения -1 и 0. Охватим ли мы таким образом все точки прямой A? Вернемся ли мы в ту же точку, откуда начали свой путь?)

СНОВА ТЕЛЕФОН

Выше мы выяснили, что по однородным координатам (X, Y, Z) точки всегда можно найти координаты (x, y), кроме того случая, когда $Z=0$.

Дальше мы можем пойти по одному из двух путей. Если мы решим, что ярлык (x, y) для нас важнее, то нам придется признать, что числа (X, Y, Z) определяют положение точки во всех случаях, кроме случая $Z=0$. Но допустим, у нас нет причин, которые заставили бы нас отдать предпочтение координатам (x, y); тогда мы можем с таким же успехом сказать, что числа (X, Y, Z) *всегда* определяют точку, и считать *недостатком* системы (x, y) отсутствие в ней ярлыков для точек, у которых $Z=0$.

Если бы нам нужно было все это объяснить нашему другу-ангелу по телефону, то мы бы увидели, что ему все равно, по какому из двух путей пойти. Все зависело бы от того, как мы изложим вопрос. Допустим, мы предлагаем нашему подрядчику на небесах следующие условия построения вселенной:

1. Вселенная должна состоять из точек, определяемых числами (X, Y, Z).
2. Важны лишь отношения этих чисел. Точка с ярлыком (kX, kY, kZ) та же, что и с (X, Y, Z), каково бы ни было число k.
3. Любому ярлыку (X, Y, Z) соответствует точка в этой вселенной; исключение составляют лишь случаи, когда $Z=0$.
4. Все прочие детали мы оставляем на усмотрение архитектора.

Как вам известно, архитекторы редко соглашаются принимать условия. Можно представить возражения нашего ангела: «Мне совсем не нравится ваш пункт 3. Наша фирма славится сооружением вселенных, отличающихся математическим изяществом. Не будете ли вы более сговорчивыми в вопросе, касающемся Z? Если вы хотите исключить случай $(0, 0, 0)$, я склонен согласиться с вами, но я не вижу ничего плохого в пропорции вида $6:3:0$».

Если бы мы полностью уступили, вселенная была бы построена на основе проективной геометрии. Если бы ангел согласился с нашими оговорками, то получилась бы вселенная с обычной евклидовой геометрией. Но если было бы принято компромиссное решение, по которому все точки, имеющие $Z=0$, были бы снабжены объявлением: «Эта точка находится за пределами вселенной, в бесконечности», то получилась бы (я думаю) евклидова геометрия, рассматриваемая как частный случай проективной.

ПРЯМЫЕ И ТОЧКИ

В предыдущей главе мы видели, что существуют проективные теоремы — теоремы, использующие только понятие *точки* и *прямой*, точек, *лежащих на* прямых, и прямых, *проходящих через* точки. Мы должны теперь определить, что́ мы понимаем под словом «прямая» в той вселенной, которую мы конструируем.

Как определяется прямая в обычной аналитической геометрии? Хорошо известно, что уравнение $ax+by+c=0$ в декартовой системе координат x, y всегда определяет прямую и что любая прямая может быть представлена подобным образом. (Более известная форма записи $y=mx+c$ оказывается непригодной для вертикальных прямых.) Если в первом уравнении мы заменим x на $\frac{Y}{Z}$, а y на $\frac{X}{Z}$ и умножим все уравнение на Z, то получим уравнение

$$aX + bY + cZ = 0.$$

В основе этого рассуждения лежит наша обычная геометрия. Поскольку новая вселенная не обязана подчиняться законам старой, вышеприведенное рассуждение теперь ничего не *доказывает*. Но она позволяет кое-что *предположить*. Если новая вселенная должна оказаться полезной для понимания старой, то она должна быть по крайней мере подобна старой. Поэтому мы решаем воспользоваться приведенным выше уравнением и принимаем следующее *определение прямой*: все точки (XYZ), удовлетворяющие уравнению $aX+bY+cZ=0$, образуют прямую. Точки, которые удовлетворяют этому уравнению, будут называться также «лежащими на прямой», а прямая будет называться «проходящей через эти точки».

Все это нетрудно передать по телефону, и наш ангел поймет, что мы имеем в виду, когда говорим о точках на прямой. Мы, конечно, не можем определить прямую как кратчайшее расстояние между двумя точками, потому что мы конструируем проективную геометрию, в которой не существует такого понятия, как *расстояние*. Обитатели этой вселенной были бы совершенно не способны придать какое-либо значение этому слову. Причина, по которой мы отбрасываем понятие расстояния, та, что, как мы видели во второй главе, чем *меньше* данных включает теория, тем проще и мощнее она становится.

Предположим теперь, что у нас есть две прямые, скажем,

$$aX+bY+cZ=0, \tag{1}$$

$$pX+qY+rZ=0. \tag{2}$$

Пересекутся ли эти прямые? Другими словами, имеют ли они общую точку? Нетрудно видеть, что если вы возьмете $X=br-cq$, $Y=cp-ax$, $Z=aq-bp$, то эти значения удовлетворяют уравнениям (1) и (2) [1]. Следовательно, всегда имеется точка, где две прямые пересекаются. В проективной геометрии не существует понятия параллельных прямых.

---

[1] По предложению ангела мы согласились, что группу $X=0$, $Y=0$, $Z=0$ не следует рассматривать как точку; в противном случае эта точка лежала бы на каждой прямой, что было бы неудобно. Затруднение могло бы возникнуть в том случае, если приведенное в тексте решение оказалось бы (0, 0, 0). Можно показать, что это происходит лишь в том случае, когда прямые (1) и (2) совпадают; тогда, разумеется, все точки прямой (1) лежат также на прямой (2).

Как же в евклидовой геометрии прямым удается быть параллельными? Возьмем в обычной системе координат (x, y) две параллельные прямые, скажем $y = x + 1$ и $y = x + 2$. Если мы переведем их в систему координат (XYZ) по способу $x = \frac{X}{Z}$, $y = \frac{Y}{Z}$, то они примут вид $X - Y + Z = 0$ и $X - Y + 2Z = 0$. Эти прямые пересекаются в точке $X = 1$, $Y = 1$, $Z = 0$, и именно потому, что $Z = 0$, мы не можем найти точку (x, y), принадлежащую одновременно двум прямым. Значит, если мы признаем несуществующими все точки, у которых $Z = 0$, и, скажем, что две прямые, пересекающиеся в точке с $Z = 0$, следует рассматривать как непересекающиеся, то от проективной геометрии мы снова вернемся к геометрии с параллельными прямыми.

В действительности мы делаем это всякий раз, когда смотрим на картину. Мы тогда видим прямые, которые нам кажутся пересекающимися на линии горизонта, и автоматически отбрасываем точки горизонта, потому что они не представляют «реальных» точек.

Между прочим, стоит отметить, что точки, которые мы признаем несуществующими в нашей вселенной, представляются уравнением $Z=0$; это уравнение линейно и является уравнением прямой. Поэтому мы и говорим о «прямой в бесконечности». Геометрически это хорошо согласуется с тем фактом, что горизонт кажется прямой линией. С точки зрения проективной геометрии прямая в бесконечности ничем не отличается от любой другой прямой. На листе бумаги вы можете начертить любую прямую и сказать: «Эта линия представляет горизонт». Конечно, вам придется повернуть рисунок так, чтобы начерченная прямая была горизонтальной.

Определения проективной геометрии очень просты. Любые три числа X, Y, Z определяют точку при условии, что они не равны нулю одновременно[1]. Уравнение $aX+bY+cZ=0$ определяет прямую (при условии, что числа a, b, c не все равны нулю). Любые две точки можно соединить прямой; любые две прямые пересекаются в одной точке.

Система очень проста. Всякому, кто привыкнет к ней, наша обычная геометрия представилась бы в следующем виде. Из совокупности всех прямых линий выделяется одна. Хотя она ничем не отличается от всех остальных, она имеет особое название. Она называется «прямая в бесконечности». Особое название дается и таким прямым, точка пересечения которых лежит на этой прямой. Они называются параллельными. С точками выделенной особой прямой обращаются так, как будто они вообще не существуют. Поэтому-то и считается, что параллельные прямые не пересекаются.

---

[1] Причем, как уже говорилось, (X, Y, Z) и (kX, kY, kZ) определяют одну и ту же точку, если $k \neq 0$. Именно этим обусловлено название «однородные координаты».— *Прим. ред.*

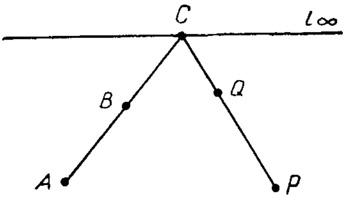


Рис. 70.

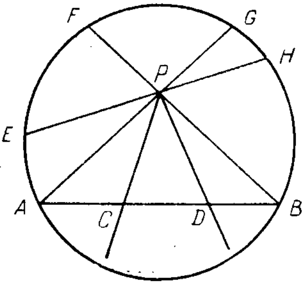


Рис. 71.

Итак, евклидова геометрия может быть получена из проективной, если исключить точки некоторой прямой (рис. 70). Прямая PQ считается параллельной AB только в том случае, если точка их пересечения является исключенной точкой. Все исключенные точки лежат на особой прямой — прямой в бесконечности (обозначим ее для краткости l_∞). Единственной исключенной точкой прямой AB является C — точка, где $l\infty$ пересекает AB. Следовательно, если прямая PQ параллельна AB (пересекается с этой прямой в исключенной точке), она должна пересекать AB в точке C. Соединяя точки C и P, получаем прямую, проходящую через P и параллельную AB. Ясно, что такое построение дает одну и только одну прямую.

ДРУГИЕ ГЕОМЕТРИИ

В главе шестой мы видели, что возможность построения геометрий, отличных от евклидовой, является не чисто теоретическим предположением, но может иметь важное значение в физике. Процедура, которую мы только что описали, открывает путь к построению других геометрических систем. Почему мы должны были объявить исключенными точки некоторой прямой? Почему бы нам не выбрать какую-нибудь кривую или область? На рис. 71 представлен обыкновенный круг; мы объявляем исключенными все точки, лежащие за пределами этого круга и на его границе. Эту граничную окружность мы можем, если угодно, назвать «окружностью в бесконечности»; математики обычно называют ее «абсолютом». Две прямые считаются пересекающимися только в том случае, когда точка их пересечения лежит *внутри* круга. Так, прямые PC и PD считаются пересекающимися с прямой AB, а прямая PA считается параллельной прямой AB, поскольку она пересекается с ней на границе круга, т. е. «в бесконечности». Будем вращать прямую PA по часовой стрелке. В положении PE она еще не пересекает прямую AB. Если продолжать вращение в том же направлении, то в положении PF наша прямая снова окажется параллельной AB, поскольку она пересекает AB в точке B на границе круга. При дальнейшем вращении по часовой стрелке прямая будет пересекать AB в «реальных» точках. Таким образом, имеется целый пучок прямых, проходящих через точку P и параллельных AB.

Такое поведение прямых несомненно напомнит вам вселенную Пуанкаре, о которой шла речь в главе шестой. Действительно, это слегка замаскированная вселенная Пуанкаре. Проектируя вселенную Пуанкаре на соответствующим образом расположенную сферу и затем обратно на плоскость, мы могли бы получить только что рассмотренную геометрию.

ТРЕУГОЛЬНИК ЦВЕТОВ

На рис. 72 представлен (в перспективе) кусок прямоугольной координатной сетки. Точка O, как обычно, — начало координат, OA, OB — оси координат, AB — линия горизонта. В обычной системе координат OA была бы прямой $y = 0$, OB — прямой $x = 0$. В однородных координатах X, Y, Z уравнения практически те же самые. Поскольку $y = \frac{Y}{Z}$, то уравнение OA будет $Y = 0$. Аналогично $X = 0$ — уравнение прямой OB.

Линия горизонта AB представляет прямую в бесконечности, уравнение которой есть $Z=0$. Фигура, таким образом, обладает трехсторонней симметрией. Обычно координатная сетка представляется нам имеющей двустороннюю симметрию. Имеются две оси: ось абсцисс и ось ординат — каждая ничем не лучше и не хуже другой. Но на нашем рисунке изображены три равноправные прямые: прямая OB, помеченная $X=0$, прямая OA, помеченная $Y=0$, и AB, помеченная $Z=0$.

Таким образом, треугольник OAB выступает как базисный. O есть точка $(0,\ 0,\ 1)$, A — точка $(1,\ 0,\ 0)$, B — точка $(0,\ 1,\ 0)$. Вы можете сверить, насколько это согласуется с нашими предыдущими представлениями, подставив эти значения в уравнения $x = \frac{X}{Z}$, $y = \frac{Y}{Z}$. Что касается точки O, здесь затруднений не возникает, поскольку $x = 0$, $y = 0$. Для точки A мы находим: $x = \infty$, $y = 0$, так что она должна находиться на оси Ox на бесконечно большом расстоянии от начала координат; но она там и находится. Подобным же образом находим для точки B: $x = 0$, $y = \infty$, т.е. точка B лежит, как и требовалось, на оси ординат на бесконечном расстоянии от O.

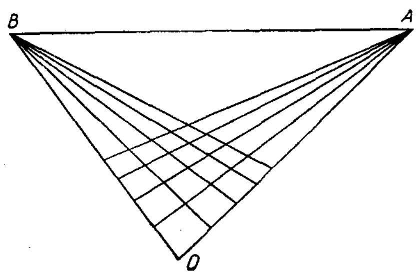


Рис. 72.

В обычной системе координат прямая, проходящая через начало, представляется уравнением $y=mx$. В однородных же координатах прямая, проходящая через O, определяется уравнением $Y=mX$. А как выглядят уравнения прямых, проходящих через точку A? Эти прямые параллельны оси абсцисс, т. е. для них $y=0$. В однородных координатах это уравнение превращается в $Y=cZ$. Аналогично уравнения прямых, проходящих через точку B, имеют вид $X=kZ$. И снова здесь наблюдается симметрия; прямые, проходящие через точки A, B и O, выступают в одинаковых ролях: один и тот же тип уравнения описывает любую из них.

Поскольку точка A имеет координаты (1, 0, 0), B — координаты (0, 1, 0), а O — координаты (0, 0, 1), мы можем рассматривать любую точку (X, Y, Z) как «смесь» точек A, B, O. Нам стоит только взять X раз координаты точки A, Y раз координаты точки B и Z раз координаты точки O, а затем сложить их. Слово «смесь» здесь употребляется в том же самом смысле, что и раньше, когда «смешивание» точек сравнивалось с приготовлением чая из разных сортов.

Та же самая идея смешивания используется и в «треугольнике цветов». Процесс цветной печати основывается на том, что любой цвет может быть получен смешиванием в определенных пропорциях трех основных цветов — красного, синего и желтого. Эффект смешивания можно проиллюстрировать на треугольнике цветов (рис. 73). Предположим, мы нанесем в точке A чистую красную краску, в точке B — чистую синюю и в точке O — чистую желтую. В точке D — середине отрезка AB — будет нанесена краска, полученная смешиванием красной и синей в равных количествах. В точке E окажется смесь равных количеств желтой и синей красок, а в F — красной и желтой. В точке G — центре треугольника мы увидим смесь красной, синей и желтой красок в равных количествах.

Называя G точкой (1, 1, 1), D точкой (1, 1, 0) и т. д., мы видим, что координаты X, Y, Z в любой точке указывают ту пропорцию, в которой смешивались в этом месте три основных цвета. Это еще раз показывает, что существенны только отношения величин X, Y, Z. Точка (2, 2, 2), например, расположена в том месте, где 2 литра красной краски смешаны с 2 литрами синей и 2 литрами желтой; но ведь тот же самый цвет получается, если взять краски в количестве одного литра каждую (2, 2, 2) — это та же точка, что и (1, 1, 1).

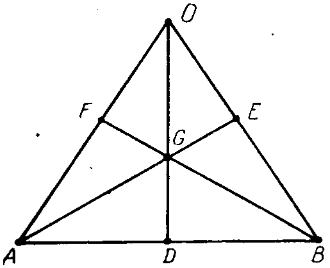


Рис. 73.

Треугольные диаграммы подобного типа используются химиками, когда им нужно знать эффект смешивания трех ингредиентов в различных пропорциях.

Всякий, кто знаком с теоретической механикой, обнаружит еще одно применение такой диаграммы. Точка (X, Y, Z) есть центр тяжести системы тел, массы которых, равные X, Y, Z, сосредоточены соответственно в точках A, B, O.

Мы совершенно отошли от того пути, на котором были впервые введены координаты X, Y, Z при помощи координатной сетки. Теперь мы можем прямо начать с *любого* треугольника, приняв его за базисный. Это дает нам гораздо большую свободу. Если мы решаем некоторую задачу о треугольнике, то нам нет нужды вводить координатную систему с осями, образующими прямой угол. Мы можем принять сам треугольник за основу системы координат.

И все же мы еще далеки от того, чтобы представить однородные координаты в самом общем виде. Идея центра тяжести — это тоже *метрическая* идея, она опирается на понятие длины. Однако можно ввести однородные координаты и чисто проективным способом, приняв любые четыре точки плоскости за (1, 0, 0), (0, 1, 0), (0, 0, 1) и (1, 1, 1), а затем начертив сеть линий, покрывающую всю плоскость, причем каждая точка сети получает свое наименование (X, Y, Z) без всякого обращения к понятию длины. Подробности этого построения можно найти в книгах по проективной геометрии (получающееся при этом множество точек получило название «сеть Мёбиуса»).

МНИМЫЕ ТОЧКИ

В нашей инструкции по построению вселенной первое правило гласит: «Вселенная состоит из точек, определяемых числами (X, Y, Z)». Здесь ничего не говорится о природе этих чисел. До сих пор, не подчеркивая этого, мы считали числа *действительными*. Но не существует никаких особых причин, чтобы они были действительными. Все наши геометрические результаты получены с помощью алгебраических вычислений. Мы можем продолжать их при условии, что числа, используемые в качестве координат X, Y, Z, подчиняются всем законам алгебры. Хорошо известно, что *комплексные числа*, т. е. числа вида $a + bi$, где $i = \sqrt{-1}$, подчиняются всем законам алгебры. Следовательно, мы можем построить вселенную, где каждая точка будет иметь координаты X, Y, Z, причем числа X, Y, Z будут комплексными. Описать эту вселенную математически было бы нисколько не сложнее, чем при действительных числах X, Y, Z, потому что комплексные числа ведут себя при алгебраических вычислениях в точности так же, как действительные.

Но, конечно, геометрия этой вселенной сильно отличалась бы от той, к которой мы привыкли. В области комплексных чисел любое уравнение имеет корень. На языке геометрии это означает, что любые две кривые пересекаются! Но ведь в обычной геометрии нам нетрудно нарисовать, например, два круга, которые не пересекаются.

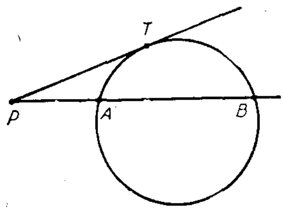


Рис. 74.

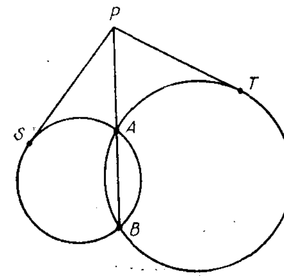


Рис. 75.

Мы можем, однако, преодолеть это расхождение, добавив к нашим правилам построения вселенной условие 1а: для существ, населяющих эту вселенную, постижимы только те точки, координаты которых X, Y, Z — действительные числа.

Теперь мы имеем два вида исключения для точек: 1) точки с $Z=0$, объявленные «точками в бесконечности»,— их не могут достичь обитатели вселенной; 2) точки с комплексными значениями X, Y, Z — их вообще нельзя воспринять[1]. Они называются мнимыми точками.

Эти два вида исключения существенно различны. Две кривые, которые пересекаются в бесконечности, называются параллельными, но кривые, пересекающиеся в мнимых точках, вовсе не называются параллельными. Обитатели вселенной просто говорят: «Эти кривые не пересекаются».

Итак, мы вводим точки с мнимыми координатами, чтобы затем снова исключить их? Чего же мы собираемся достичь при помощи этой курьезной процедуры?

Ее полезность можно показать на следующем примере. В евклидовой геометрии имеется известная теорема: $PT^2=PA\cdot PB$ (рис. 74).

Теперь представьте, что у нас есть две окружности, каждая из которых проходит через точки A и B (рис. 75). Из произвольной точки P, лежащей на прямой AB, проводим касательные PS и PT к окружностям. Согласно приведенной выше теореме, $PS^2=PA\cdot PB=PT^2$; поэтому $PS=PT$. Этот результат известен еще из курса школьной геометрии: касательные к двум окружностям, проведенные из точки P, лежащей на прямой AB, равны.

Доказательство этой теоремы несложно, но вот что досадно: в курсе школьной геометрии оно возможно лишь в том случае, когда окружности пересекаются. Между тем, теорема: «Существует

[1] Строго говоря, необходимо только, чтобы действительными были отношения чисел $X:Y:Z$. Точка (i, i, i) — та же самая, что точка с координатами $(1, 1, 1)$, и она считается доступной для обитателей вселенной.

такая прямая, что касательные к двум окружностям, приведенные из любой ее точки, равны» — верна и для случая непересекающихся кругов. Но доказывать это приходится уже другим способом!

Теперь посмотрим на задачу с алгебраической стороны. Возьмем две непересекающиеся окружности, скажем, одну с центром (0,20) и радиусом 16, а другую — с центром в точке (0, —15) и радиусом 9. Уравнения этих окружностей:

$$x^2+(y-20)^2=16^2 \text{ и } x^2+(y+15)^2=9^2.$$

Эти две окружности в действительности не пересекаются; однако предположим все-таки, что они пересекаются. Тогда было бы естественно попытаться найти точки пересечения этих окружностей A и B и уравнение прямой AB, на которой должна была бы лежать точка P. Чтобы найти точки A и B, нам нужно решить эти уравнения совместно. Решение получилось бы в виде $x=12i$, $y=0$ и $x=-12i$, $y=0$. Вы можете легко проверить, что эти значения удовлетворяют обоим уравнениям, т. е. точки $(12i, 0)$ и $(-12i, 0)$ принадлежат обеим окружностям. Следовательно, это и должны быть искомые точки A и B. Обе они лежат на прямой $y=0$. Итак, уравнение прямой AB будет $y=0$.

Но $y=0$ — это самая настоящая реальная прямая. Более того, если вы начертите обе окружности в прямоугольной системе координат и возьмете любую точку на оси y, то касательные, проведенные из этой точки к обеим окружностям, окажутся равными. А если вы достаточно хорошо знакомы с аналитической геометрией, то вы сможете даже доказать этот результат с помощью вычислений.

Таким образом, проделано следующее: мы вели вычисления так как если бы окружности пересекались в реальных точках; мы продолжали вычисления даже тогда, когда нам встречалось число $\sqrt{-1}$; и мы пришли к правильному результату, в котором число $\sqrt{-1}$ отсутствует!

Но это, видимо, означает, что нам не нужно больше беспокоиться, пересекаются ли окружности в реальных точках или нет, нужно проводить одни и те же вычисления в любом случае. Применяя этот принцип в геометрии, мы избавляем себя от необходимости рассматривать бесчисленное количество вариантов, возникающих в некоторых случаях. Например, в задаче о трех кругах[1] могут встретиться следующие случаи: все три окружности пересекаются; они не пересекаются; пересекаются только какие-нибудь две окружности из трех или одна окружность пересекает две другие, которые между собой не пересекаются. Можно избежать всех этих подробностей, если допустить существование мнимых точек пересечения.

---

[1] Так называется следующая задача: построить окружность, касающуюся трех данных окружностей.— *Прим. ред.*

И пусть вас не волнуют философские размышления о том, «действительно ли существуют» эти мнимые точки. Математика имеет дело с закономерностями, а не с предметами. Если мы сможем показать, что вселенная, построенная по нашим правилам (с дополнительным условием 1а), и евклидова плоскость обладают одними и теми же закономерностями, то это все, что нам нужно было сделать. Существует ли в действительности где-нибудь такая вселенная,— это нисколько не отразится на логике метода. Важно лишь, что она могла бы существовать без логических противоречий.

ПЕРЕСЕЧЕНИЕ ОКРУЖНОСТИ С БЕСКОНЕЧНО УДАЛЕННОЙ ПРЯМОЙ [1]

В предыдущем разделе, рассматривая задачу о двух окружностях, мы использовали обычную систему координат x, y. Поскольку бесконечно удаленная прямая не участвовала в решении этой задачи, система x, y прекрасно служила нашим целям. Единственным новшеством было введение комплексных чисел.

Однако сейчас мы собираемся спросить: где окружность пересекает бесконечно удаленную прямую? Поскольку в дело вступает бесконечность, нам придется использовать координаты X, Y, Z. А поскольку в действительности окружности не заходят в бесконечность, точки пересечения должны быть мнимыми, значит, в ответе появятся комплексные числа.

В обычной системе координат x, y окружность с центром в точке (a, b) и радиусом r представляется уравнением

$$(x-a)^2+(y-b^2)=r^2.$$

В однородных координатах, введенных с помощью формул

$$x=\frac{X}{Z} \text{ и } y=\frac{Y}{Z},$$

уравнение окружности принимает вид

$$(X-aZ)^2+(Y-bZ)^2=r^2Z^2.$$

Уравнение бесконечно удаленной прямой есть $Z=0$. Для нахождения точек пересечения окружности с этой прямой подставляем значение $Z=0$ в уравнение окружности. Получаем $X^2+Y^2=0$. Отсюда $Y^2=-X^2$. Извлекая квадратный корень, находим $Y=iX$ или $Y=-iX$. Мы можем выбрать любое значение для X, поскольку важно лишь отношение $X:Y:Z$. Возьмем $x=1$. Таким образом, мы находим, что точки пересечения имеют коорди-

[1] В интересах ясности изложения в этом разделе мы изменили термин «прямая в бесконечности» на «бесконечно удаленная прямая».— *Прим. ред.*

наты (1, i, 0) и (1, $-i$, 0). Эти точки обычно обозначают буквами I и J.

Замечательно, что точки пересечения нисколько не зависят от величин a, b, r, которые отличают одну окружность от другой. Другими словами, независимо от того, где расположен центр окружности или каков ее радиус, окружность пересекает прямую $Z=0$ всегда в одних и тех же точках I и J. Это очень неожиданный результат. Он дает нам возможность дать новое определение понятия «окружность».

Прямая линия всегда задается уравнением вида $aX+bY+cZ=0$. Каждый член уравнения содержит только одну из величин X, Y, Z и притом в первой степени; поэтому уравнение и называется уравнением *первой степени*. Естественно после изучения прямых перейти к рассмотрению кривых, определяемых уравнением *второй степени*. Любая такая кривая называется *коническим сечением*. Ее уравнение имеет вид:

$$aX^2+bY^2+cZ^2+2fYZ+2gZX+2hXY=0.$$

Нетрудно показать, что коническое сечение, проходящее через точки I и J, должно быть окружностью.

Итак, мы пришли к совершенно новому определению окружности: окружность — это коническое сечение, проходящее через две точки I и J. Это определение, хотя, возможно, и кажется странным, на самом деле вполне строгое и в математическом отношении гораздо проще, чем туманные понятия евклидовой геометрии. Оно является исходным пунктом современного развития геометрии окружностей.

Глава двенадцатая

О ПРЕОБРАЗОВАНИЯХ

Искусство рассуждения состоит в том, чтобы сразу подойти к вопросу с правильной стороны, ухватиться за ряд общих идей, которые проясняют целое, и настойчиво группировать вокруг них все второстепенные факты. Нельзя стать хорошим аналитиком, пока в результате постоянной практики не научишься понимать, насколько важно схватить главную идею и держать ее мертвой хваткой.

А. Н. Уайтхед,
Президентское послание Лондонскому отделению Математической ассоциации, 1914 г.

В главе десятой мы рассматривали вопрос о том, как одна проблема может оказаться замаскированной формой другой. Такое изменение внешней формы называется *преобразованием*. Ясно, что преобразования умножают наши знания и экономят усилия.

Рассмотрим вначале пару самых элементарных примеров преобразований. Здесь мы не получим каких-либо неожиданных результатов, а просто покажем значение преобразований в их самой простой и ясной форме, без всяких усложнений. Позже мы перейдем к другой крайности, чтобы показать ту грандиозную и неожиданную силу, которой обладают преобразования.

Рассмотрим следующие две задачи: (1) найти корень квадратный из 2; (2) найти корень квадратный из 200.

Ответ к первой задаче очень прост, нужно только заглянуть в таблицу квадратных корней и найти, что $\sqrt{2}$ приблизительно равен 1,4142. Если мы попытаемся отыскать в таблице корень квадратный из 200, то мы не найдем его, потому что, как правило, в таблице приведены только значения квадратных корней из чисел от 1 до 100. Причина, конечно, заключается в том, что $\sqrt{200}$ — это просто $10\sqrt{2}$. Умножение на 10 — это такая простая операция, что было бы неразумно печатать отдельные таблицы корней для чисел, превосходящих 100. Принцип нахождения $\sqrt{200}$ может быть использован для нахождения квадратного корня из любого числа, большого или малого. Таким образом, благодаря этому принципу, возможно неограниченное распространение таблицы. Итак, наряду с $\sqrt{2}$ мы имеем ряд замаскированных форм: $\sqrt{200}$, $\sqrt{20\,000}$, ..., $\sqrt{0{,}02}$, $\sqrt{0{,}002}$, ... и т. д. до бесконечности.

Принцип отыскания $\sqrt{200}$ может быть алгебраически выражен в следующем виде: чтобы решить уравнение $x^2 = 200$ (2), мы применяем *преобразование* $x = 10\,y$, которое сводит задачу к решению уравнения (1): $y^2 = 2$.

Преобразование может быть полезным, если его применение несложно. Здесь, в нашем примере, нам пришлось лишь умножить на 10, что делается весьма легко. Другие преобразования, естественно, более сложны, чем это, но, вообще говоря, мы обычно удовлетворены, если преобразование *проще, чем сама преобразуемая задача.* В задачах на интегрирование мы считаем простым любое преобразование, которое может быть выполнено чисто алгебраически — таким, например, является проективное преобразование, примененное для преобразования подынтегральной функции в конце главы 10.

ПРЕОБРАЗОВАНИЯ И УРАВНЕНИЯ

Только что рассмотренные примеры легко обобщить. Можно ли найти преобразование, которое приводило бы данное уравнение к более простому виду?

В школе мы изучаем различные методы решения квадратных уравнений: разложение на множители, дополнение до полного квадрата или просто решение по формуле. Обо всех этих методах нам *рассказывают.* Давайте подойдем к задаче решения квадратных уравнений так, как если бы мы были первыми людьми, которые встретились с этой задачей.

Прежде всего посмотрим, что нам потребовалось бы для составления таблиц решений квадратных уравнений, если бы мы просто решили составить таблицу всех решений. Нетрудно составить таблицу, которая содержала бы тысячу значений — вспомните обычную таблицу логарифмов, где приводятся значения логарифмов всех чисел от 1,00 до 9,99; девятьсот значений обычно занимают пару страниц. Но если мы захотим составить таблицу для функции двух переменных, каждая из которых может принимать тысячу значений, то в такой таблице должен содержаться миллион значений. Например, если мы имеем две переменные a и b, каждая из которых принимает значения от 0,00 до 9,99, то первый лист нашей таблицы должен содержать ответы для случая, когда $a=0{,}00$, а b принимает значения от 0,00 до 9,99, второй — для случая $a=0{,}01$ и для b в том же диапазоне и т. д. до тысячного листа, где a равно 9,99, а b изменяется от 0,00 до 9,99.

Таким образом, таблица для двух переменных заняла бы уже не один лист, а целую книгу из тысячи листов. Аналогично, таблица для трех переменных потребовала бы тысячи книг по тысяче листов в каждой. Все это хорошо известно составителям таблиц.

Квадратное уравнение $ax^2+bx+c=0$ содержит три постоянных a, b, c. На первый взгляд кажется, что таблицы для записи всех решений этого уравнения потребовали бы грандиозной библиотеки, о которой говорилось выше. Некоторое облегчение приносит тот факт, что мы можем разделить уравнение на a и таким образом сделать коэффициент при x^2 равным 1. Это приводит нас к уравнению вида $x^2+px+q=0$. Но даже и в этом случае мы будем иметь две величины, p и q, и том на тысячу листов все еще кажется необходимым.

Можно ли улучшить положение при помощи преобразования, то есть можем ли мы упростить уравнение, заменив x некоторым выражением, содержащим новую переменную y?

Возьмем, например, уравнение $x^2-3x-5=0$. Как уже упоминалось, наше преобразование должно быть простым. Поскольку исходная задача — это уравнение второй степени, то предполагается, что преобразование должно быть еще проще, т. е. должно быть линейным. Мы можем попытаться использовать общее линейное выражение вида $x=my+c$. Однако оказывается, что совершенно достаточно взять просто $x=y+c$. Подставляя это выражение в исходное уравнение, получаем после умножения и приведения подобных членов: $y^2+(2c-3)y+c^2-3c-5=0$. Здесь нужно сделать простейшую вещь — преобразовать коэффициент при y, а именно $2c-3$. Если мы выберем c равным $1\frac{1}{2}$, то этот коэффициент будет равен нулю.

Итак, мы находим, что уравнение $x^2-3x-5=0$ упрощается с помощью преобразования $x=y+1\frac{1}{2}$, которое приводит его к виду $y^2-7\frac{1}{4}=0$.

Это последнее уравнение равносильно уравнению $y^2=7{,}25$, решение которого можно найти непосредственно в таблице квадратных корней. То обстоятельство, что в приведенном уравнении встречается дробь $\frac{1}{4}$, в то время как в исходном уравнении дробей не было, несущественно: найти квадратный корень из 7,25 так же легко, как квадратный корень из 7. Затем мы можем получить решение исходного уравнения добавлением к ответу числа $1\frac{1}{2}$.

В действительности любое квадратное уравнение можно с помощью подходящего преобразования привести к такому виду, что для его решения останется только заглянуть в таблицу квадратных корней.

Таким образом, вместо фантастического собрания томов, которое казалось необходимым на первый взгляд, мы можем вполне удовлетвориться обычной таблицей квадратных корней для того, чтобы решить любое квадратное уравнение.

Этот частный пример иллюстрирует ту роль, которую играют преобразования, и экономию, к которой они приводят. Пример элементарен, но заложенный в нем принцип с таким же успехом применим во многих серьезных разделах математики.

ПРОСТОЙ ПЕРЕНОС ГРАФИКА

При изучении какой-либо кривой часто возникает вопрос о возможности перевода ее в другое положение. Например, хорошо известно, что график функции $y=x^2$ есть U-образная кривая, вроде изображенной на рис. 76, где A — начало координат, P — некоторая произвольная точка кривой, Q — точка на оси абсцисс, лежащая под точкой P. Поскольку $PQ=y$, а $AQ=x$, то справедливо соотношение $PQ=AQ^2$.

Предположим, что мы начертили эту кривую на кальке и передвинули кальку так, что точка A совпала с точкой (3,2). Эта ситуация изображена на рис. 77. Как будет выглядеть уравнение кривой в ее новом положении?

Поскольку кривая не изменилась, то соотношение $PQ=AQ^2$ все еще верно, но было бы неправильным сказать, что AQ равно x, а PQ равно y. Расстояние точки P от нового начала координат O равно не AQ, а $3+AQ$, откуда $AQ=x-3$. Аналогично получаем, что $PQ=y-2$. Следовательно, соотношение $PQ=AQ^2$, переведенное на язык алгебры, принимает вид $y-2=(x-3)^2$, что после упрощения дает $y=x^2-6x+11$.

Отсюда следует, что если у вас есть график функции $y=x^2$, а вам нужно начертить график функции $y=x^2-6x+11$, то для этого вам не нужно делать никаких вычислений. Новая кривая может быть получена из старой перемещением на 3 единицы на восток и на 2 единицы на север.

Нетрудно показать, что график функции вида $y=x^2+px+q$ есть просто график $y=x^2$, смещенный на некоторое расстояние относительно его исходного положения.

Если вам известно все относительно графика $y=x^2$, то вам известно все и о графике вида $y=x^2+px+q$.

КОНФОРМНЫЕ ПРЕОБРАЗОВАНИЯ

В главе первой, в разделе, озаглавленном «Какие закономерности любит природа», приводился перечень более десятка вопросов практической важности, которые имеют определенную связь с числом $\sqrt{-1}$.

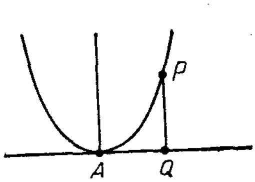


Рис. 76.

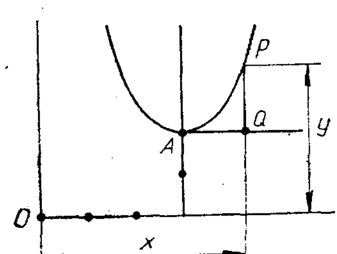


Рис. 77.

Некоторые задачи, встречающиеся в этой области, обладают тем свойством, что их можно преобразовать в целое множество других задач, таких, что если мы знаем решение одной задачи, мы можем сразу же получить решения всего семейства задач.

Используемые при этом преобразования называют *конформными*. Прежде всего, я попытаюсь объяснить, что такое конформное преобразование, а затем приведу пример, демонстрирующий мощь этого метода.

Каждой функции $f(z)$ — или, во всяком случае, любой «разумной» функции — соответствует конформное преобразование. Можно взять функцию $z^3 - 7z + 2$, или e^z, или $\log z$, или $\sqrt{z+3}$, или еще какую-нибудь функцию; вид функции не играет роли. Каждой из этих функций соответствует преобразование, и все преобразования различны. Благодаря большой свободе выбора мы можем преобразовать одну задачу во множество других.

Для того чтобы показать, как выполняется такое преобразование, выберем очень простую функцию z^2 и покажем, что делает соответствующее ей преобразование с какой-либо фигурой.

Рассмотрим четыре квадрата (рис. 78), образованные отрезками прямых $x=1$, $x=2$, $x=3$, $y=1$, $y=2$, $y=3$ в обычной декартовой системе координат. Преобразование выполняется в три этапа:

1. Пишем комплексное число, соответствующее каждой точке чертежа. 2. Возводим это число в квадрат. (Это потому, что мы выбрали в качестве исходной функцию z^2. Если бы мы выбрали какую-нибудь другую функцию $f(z)$, то нам нужно было бы вычислить значение этой функции для нашего комплексного числа.) 3. Строим точку, соответствующую полученному числу.

Все это очень легко выполнить. Например, если мы хотим узнать, куда в результате преобразования попадает точка (2, 3), мы должны проделать следующее:

Этап 1. Берем комплексное число, соответствующее точке (2, 3); это число есть $2+3i$. В общем случае комплексное число, соответствующее точке (x, y), имеет вид $x+iy$.

Этап 2. Возводим в квадрат число, найденное на первом этапе. Квадрат числа $2+3i$ равен $-5+12i$.

Этап 3. Находим точку, соответствующую числу $-5+12i$. Эта точка есть $(-5, 12)$. Мы используем здесь то же правило, по которому находили число на первом этапе, только в обратном направлении.

Итак, в результате преобразования точка (2, 3) попадает в положение $(-5,12)$.

Подобные вычисления проделываются для всех девяти точек, полученных пересечением исходных прямых. Находим, что точка (1,1) переходит в (0,2), точка (1,2) в $(-3,4)$, точка (1,3) в $(-8,6)$ и т. д. для всех остальных точек.

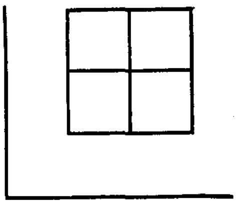


Рис. 78.

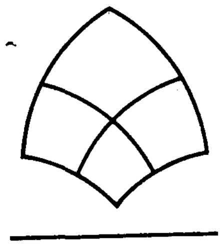


Рис. 79.

Мы наносим эти точки на координатную сетку и соединяем их кривыми линиями, соответствующими прямым исходной фигуры.

На рис. 79 показана фигура, полученная из исходной в результате преобразования z^2.

Вы видите, что преобразование действует подобно искривляющему зеркалу. Когда вы смотритесь в выпуклую часть ложки или в хорошо начищенный кувшин, вы видите создание с лицом и руками, напоминающими ваше собственное лицо и руки, однако это отнюдь не точная копия оригинала. Пропорции отчасти нарушены; но кое-что все же сохранено, и это дает вам возможность признать наблюдаемую картину отражением человека.

Конформные преобразования превращают прямые линии в кривые и изменяют расстояния. Но они сохраняют без изменения углы (на рис. 79 все кривые пересекаются под прямыми углами, как и на рис. 78), и, что особенно важно для практических приложений, они сохраняют свойство «быть решением физической задачи». Если на исходном чертеже представлены силовые линии или линии тока (для определенного класса задач), то на преобразованном чертеже будут по-прежнему представлены силовые линии или линии тока.

Замечательным примером использования этого принципа является открытая Жуковским форма крыла самолета.

Определить, каким образом поток обтекает круглое препятствие,— это простая задача с элементарным математическим решением. Линии тока изображены на рис. 80. Это есть наш исходный пункт — задача, решение которой хорошо известно. Поскольку круг является весьма простой кривой, не удивительно, что у этой задачи имеется точное, простое решение.

Практическая задача состоит в следующем: нужно найти форму воздушного потока, обтекающего крыло самолета; причем сечение крыла совсем не такое простое с геометрической точки зрения, как круг.

Жуковский открыл, что, взяв совсем простую функцию, а именно $z+\frac{1}{z}$, и применив соответствующее ей конформное преобразование,

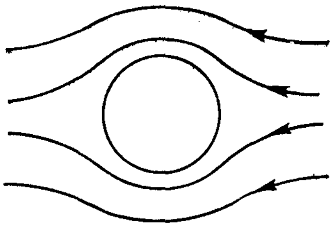


Рис. 80.

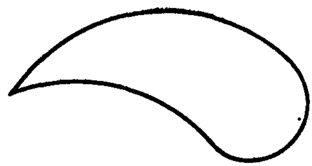


Рис. 81.

можно преобразовать форму круга к виду, который по меньшей мере напоминает сечение крыла самолета. Кривая, к которой приводится окружность в результате преобразования, изображена на рис. 81. На практике эта кривая не используется для построения крыла самолета; такое крыло не было бы эффективным. Но она полезна в теоретическом смысле, потому что с ее помощью мы получаем некоторые указания относительно формы потока, огибающего предметы такого вида. Без преобразования справиться с такой задачей было бы весьма затруднительно.

Для того чтобы решить эту задачу с помощью преобразования, мы просто преобразуем весь чертеж воздушного потока, огибающего круглый предмет. Круг переходит в фигуру, напоминающую профиль крыла. Линии потока, огибающего круглое тело, преобразуются в линии потока, огибающего крыло самолета.

На рис. 80 сделан лишь грубый набросок реальной картины. Тот, кого интересует математическое решение задачи и точная картина потока, может найти эти сведения в книгах по гидродинамике. Преобразование Жуковского и другие конформные преобразования описаны в учебниках по теории функций комплексного переменного. Для чтения этих книг необходима соответствующая математическая подготовка.

Глава тринадцатая

КОНЕЧНЫЕ АРИФМЕТИКИ И КОНЕЧНЫЕ ГЕОМЕТРИИ

Карр утверждает, что большинство австралийских племен ведет счет парами. И это для них настолько привычно, что абориген едва ли обратит внимание, если из ряда, состоящего из семи булавок, удалить две; но он немедленно заметит исчезновение одной булавки.

Тобайес Данциг,
Число — язык науки

Вопросы, рассмотренные в этой главе, выглядят очень необычными. Это их первая особенность.

Во-вторых, они играют немаловажную роль в различных разделах высшей математики.

И в-третьих, что самое неожиданное, они имеют практическое применение. Какими бы необычными и странными они ни казались, они все-таки могут оказаться полезными. В конце главы вкратце излагается одно из применений «конечной арифметики»— в агрономических исследованиях.

ОДНА НЕОБЫЧНАЯ АРИФМЕТИКА

Немногие из нас, я думаю, проверяли непосредственным вычислением, что $12\times12\times12$ действительно равно 1728. Большинство людей принимают арифметику на веру. В школах изучаются определенные правила, и они не подвергаются сомнению. Даже в наш относительно просвещенный век это относится к подавляющему большинству школьников и учителей. Людей с более критическим складом ума успокаивает общая непротиворечивость арифметики [1]. Очень часто можно произвести вычисление различными способами и прийти к одному и тому же результату. Например, если нам нужно найти $5\times(7+3)$, мы можем либо сложить 7 и 3 и, получив число 10, умножить его на 5, либо переписать выражение в виде $(5\times7)+(5\times3)$, что равно $35+15$. В обоих случаях ответ равен 50. Правила по крайней мере не приводят к противоречиям, и это дает возможность предположить, что они верны.

Арифметика, о которой мы собираемся рассказать, хотя и отличается от арифметики в привычном для нас смысле, выдержала бы испытания, подобные приведенным выше. Она построена на си-

---

[1] См. примечание к стр. 147.— *Прим. ред.*

стеме правил, очень похожих на правила обычной арифметики; они непротиворечивы[1] и всегда приводят к одному и тому же ответу.

Представьте себе, что ребенку на первых же шагах его обучения в школе нужно было бы запомнить две приведенные ниже таблицы — таблицу сложения и таблицу умножения. За каждый неправильный ответ его наказывают. И наряду с этими таблицами существуют еще и обычные правила арифметики, которых ребенок не знает.

В этой новой арифметике имеется только пять чисел: 0, 1, 2, 3, 4. В любом случае в результате операций получается одно из этих пяти чисел. Ребенку не нужно больше думать ни о десятках, ни о разрядах единиц, ни о чем-нибудь другом в этом духе.

Вот эти две таблицы, которые он должен запомнить.

Таблица сложения

	0	1	2	3	4
0	0	1	2	3	4
1	1	2	3	4	0
2	2	3	4	0	1
3	3	4	0	1	2
4	4	0	1	2	3

Таблица умножения

	0	1	2	3	4
0	0	0	0	0	0
1	0	1	2	3	4
2	0	2	4	1	3
3	0	3	1	4	2
4	0	4	3	2	1

Чтобы найти, чему равно $1+3$, ребенок смотрит в таблицу сложения. Он находит строку, где слева стоит 1, и движется по ней до пересечения со столбцом 3. В месте пересечения он читает цифру 4, что согласуется с общепринятой арифметикой. Менее обычным является результат $3+4=2$ или $2\times3=1$. Однако ребенок ничего не знает об общепринятой арифметике. Он заучивает эти таблицы, как ему велят.

Но представьте, что он захочет проверить вычисления такого же типа, как рассмотренные выше, например $3\times(2+4)$. Он может сказать, что это равно $(3\times2)+(3\times4)$[2]. В таблице умножения он находит, что $3\times2=1$ и $3\times4=2$. Итак, ответ будет $1+2=3$. Но он может также сказать: $2+4=1$, $3\times1=3$. Придя разными путями к одному и тому же результату, он убеждается в правильности того, что он заучил[3].

Возможно, вы захотите проверить другие вычисления в этой арифметике. Вы обнаружите, что, пользуясь правильно этими таблицами, вы всегда придете к одному результату, каким бы путем ни шли.

---

[1] Более того, непротиворечивость этих правил доказывается совсем элементарно.— *Прим. ред.*

[2] Конечно, лишь в предположении, что ему сообщили, что в этой арифметике верен закон *дистрибутивности*. Если же он этого не знает, то, приняв непротиворечивость на веру, он сможет таким путем *удостовериться* в дистрибутивности умножения по отношению к сложению.— *Прим. ред.*

[3] См. предыдущее примечание.— *Прим. ред.*

В действительности эта арифметика обладает рядом преимуществ по сравнению с обычной. Например, в ней отсутствуют дроби и отрицательные числа. Так, $2:3=4$, $2-3=4$. Одно число можно разделить на любое другое (исключая, конечно, деление на 0) без остатка. Любое число имеет себе обратное. Так, обратным числом для 1 является 1, для 2 — число 3, для 3 — число 2, для 4 — число 4.

Выражение $1\frac{1}{3}:2\frac{1}{4}$ упрощается следующим образом: $\frac{1}{3}=2$ (поскольку $3\times 2=1$); поэтому $1\frac{1}{3}=1+2=3$—это есть значение числителя. $2\frac{1}{4}$ находится аналогично: $\frac{1}{4}=4$ (так как $4\times 4=1$); поэтому $2\frac{1}{4}=2+4=1$. Следовательно, дробь равна $\frac{3}{4}$, т. е. 3.

На основе этой арифметики можно построить алгебру, и правила ее будут мало отличаться от правил обычной алгебры. Квадратное уравнение может быть решено способом дополнения до полного квадрата и никогда не имеет более двух решений. Например, если нам нужно решить уравнение $x^2-2x+2=0$, мы прибавляем 3 к обеим его частям и получаем $x^2-2x=3$. (Вспомним, что $2+3=0$.) Чтобы иметь слева полный квадрат, прибавим еще 1, тогда получим $x^2-2x+1=4$, т. е. $(x-1)^2=4$. 4 — это квадрат 2, но также и квадрат 3 (смотри таблицу умножения). Поэтому $x-1$ равно либо 2, либо 3, т. е. x должен быть либо 3, либо 4. Правильность полученного результата вы можете проверить, подставив ответы в исходное уравнение; конечно, вам следует быть при этом внимательными, чтобы не сбиться на обычные таблицы сложения и умножения, которые вы изучали в школе.

В такой алгебре будет ограниченное число квадратных уравнений. Коэффициент при x^2 может принимать лишь значения 1, 2, 3, 4; в случае, когда он равен 0, уравнение перестает быть квадратным. Коэффициент при x и свободный член могут принимать значения 0, 1, 2, 3, 4. Всевозможные сочетания этих коэффициентов дали бы сотню различных квадратных уравнений, так что ученик мог бы сказать, что он решил все возможные квадратные уравнения.

СТРОЕНИЕ КОНЕЧНОЙ АРИФМЕТИКИ

Откуда взялись эти таблицы? Как они построены?

Мне рассказали однажды, как в одном банке все служащие были поставлены на ноги, потому что оказалось, что не хватает ровно одного миллиона фунтов стерлингов. За достоверность этой истории я не ручаюсь; однако мне говорили, что в этом банке для расчетов применялась вычислительная машина, у которой было только шесть разрядов для целых фунтов стерлингов; поэтому максимальное количество денег, которое она могла показать, было 999 999 фунтов

19 шиллингов $11\frac{3}{4}$ пенса[1]. В кассе находилась как раз примерно такая сумма, когда кто-то внес еще пять фунтов, и машина показала результат 000 004 фунта 19 шиллингов $11\frac{3}{4}$ пенса[2] — как раз на миллион фунтов меньше, чем в действительности было в кассе.

Арифметика, которую мы рассматривали, могла бы возникнуть при применении таких вычислительных машин, у которых имеются только пять цифр: 0, 1, 2, 3, 4, расположенных по кругу, так что следующей после цифры 4 будет всегда 0. Особенность этой машины заключается в том, что прибавление 5 не меняет результата. Нигде не регистрируется, сколько оборотов сделал круг. Каждое число заменяется его остатком от деления на 5. Например, $3\times4=12$. Остаток от деления 12 на 5 равен 2, поэтому в таблице умножения появляется цифра 2 как результат умножения 3×4.

Используя только два числа 0 и 1, можно получить простейшую арифметику подобного типа, где действия производятся по правилу:

Сложение

	0	1
0	0	1
1	1	0

Умножение

	0	1
0	0	0
1	1	1

Такая арифметика возникает, когда каждое число заменяется его остатком от деления на 2. Но если хотите, мы можем заменить число 0 словом «чет», а число 1 словом «нечет»; тогда таблица сложения примет вполне понятный вид:

$$\begin{aligned} \text{чет} + \text{чет} &= \text{чет},\\ \text{чет} + \text{нечет} &= \text{нечет},\\ \text{нечет} + \text{нечет} &= \text{чет}. \end{aligned}$$

Подобным же образом можно записать и таблицу умножения.

Похоже, что именно такой арифметикой пользуется племя, о котором упоминалось в эпиграфе к этой главе. Очевидно, они считают на руках, а не на пальцах; «левая, правая, левая, правая, левая, правая...» Фактически «левая» означает «нечет», а «правая» — «чет»; забавным следствием такого способа счета является то, что они не замечают исчезновения четного количества предметов.

В главе седьмой, в разделе «Алгебра классов» мы уже пользовались двоичной системой счисления. Это был один из примеров применения двоичной арифметики к другому разделу математики.

---

[1] 1 фунт стерлингов (до перехода на десятичную систему) содержал 20 шиллингов, 1 шиллинг — 12 пенсов.— *Прим. ред.*

[2] 000 004 (т. е. просто 4) вместо 1 000 004. так как для первой цифры попросту не нашлось места.— *Прим. ред.*

КОНЕЧНЫЕ ГЕОМЕТРИИ

Сейчас мы собираемся построить еще одну вселенную. В главе одиннадцатой первое условие гласило: «Вселенная состоит из точек, определяемых тройками чисел (X, Y, Z)». Эти числа были, разумеется, числами обычной арифметики, поэтому вселенная содержала бесконечное число точек. Однако мы можем несколько изменить это условие, с тем, чтобы получить вселенную с конечным числом точек. Все, что нужно для этого сделать,— это потребовать, чтобы числа X, Y, Z принадлежали одной из рассмотренных в этой главе конечных арифметик и подчинялись законам этой системы.

Арифметика, состоящая только из чисел 0 и 1, называется «арифметикой по модулю 2». Посмотрим, какую вселенную мы получим, если условимся брать координаты X, Y, Z из арифметики по модулю 2. Как обычно, мы исключаем точку (0, 0, 0). Если бы точка (0, 0, 0) была дозволенной точкой, то в системе было бы всего 8 точек. Но поскольку она исключается, то точек в системе только 7. Вот они:

$A(1, 0, 0)$, $C(0, 0, 1)$, $E(1, 0, 1)$, $G(1, 1, 1)$.
$B(0, 1, 0)$, $D(0, 1, 1)$, $F(1, 1, 0)$.

Эти семь точек и составляют нашу вселенную.

В главе одиннадцатой мы показали, что все точки прямой могут быть получены «смешиванием» двух точек. Что мы можем получить теперь, смешивая, скажем, точки A и B? Мы очень ограничены в количестве смесей, потому что в нашем распоряжении имеются лишь числа 0 и 1. Возможны только следующие смеси:

1) 1 раз A и 0 раз B. Это дает просто A.
2) 0 раз A и 1 раз B. Это дает просто B.
3) 1 раз A и 1 раз B. Это дает $A+B$.

Складывая координаты точек A и B

$$\begin{array}{r} 1\,0\,0 \\ +\,0\,1\,0 \\ \hline 1\,1\,0 \end{array}$$

мы получаем точку F. Таким образом, $A+B=F$.

4) 0 раз A и 0 раз B. Но так мы получаем точку (0, 0, 0), которая является недозволенной. Следовательно, на прямой AB лежат лишь три точки, а именно A, B и F.

Точно так же мы могли бы показать, что на прямой, соединяющей любые две точки, лежат только три точки; на прямой AB это точки A, B и $A+B$, причем последняя есть точка F. Аналогично на AC лежат точки A, C и $A+C$; последняя есть E.

Ниже перечисляются все возможные прямые этой вселенной:

1. BDC 3. AFB 5. BGE 7. DEF
2. AEC 4. AGD 6. CGF.

К сожалению, на бумаге нельзя изобразить эту вселенную так, чтобы все прямые линии выглядели прямыми. На рис. 82 шесть

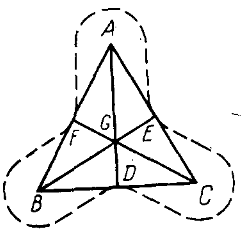


Рис. 82.

прямых представлены отрезками, но седьмая «прямая» совсем не похожа на прямую. Пунктирная кривая указывает, что точки D, E и F должны лежать на одной прямой.

Запишем уравнения перечисленных выше прямых:

1. $X=0$ 2. $Y=0$ 3. $Z=0$
4. $Y+Z=0$ 5. $X+Z=0$ 6. $X+Y=0$
7. $X+Y+Z=0$

При проверке этих уравнений вы не должны забывать, что $1+1=0$. Это следует иметь в виду все время.

Приведенный здесь чертеж является весьма неудовлетворительным, поскольку на нем точки не кажутся равноправными. Например, точка G расположена в центре чертежа, чего нельзя сказать ни об одной другой точке. Прямая DEF изображена в виде кривой, в то время как все остальные прямые изображены верно.

На самом же деле это чрезвычайно демократичная вселенная. Все точки равноправны, и все прямые также равноправны. Например, глядя на чертеж, вы можете подумать, что точка F лежит между точками A и B, поскольку $A+B=F$. Но соотношение между точками A, B и F совершенно симметрично: с тем же успехом $A=F+B$ и $B=F+A$. Каждая точка — это «смесь» двух других (это обстоятельство вы легко сможете проверить сложением координат).

В главе десятой в связи с теоремой Дезарга мы рассматривали чертежи, на которых можно было переставлять буквы, не искажая содержания теоремы. Тем же самым свойством обладает и последний чертеж. Существует 168 способов, которыми можно расставить буквы на чертеже. Сейчас буквы A, B и C расположены в вершинах треугольника; но те же вершины можно пометить любыми буквами, при условии, чтобы эти буквы не принадлежали трем точкам одной прямой. Вы можете, например, обозначить вершины буквами E, F, G, но не буквами D, E, F, поскольку D, E, F — это прямая 7.

Построенная таким образом вселенная представляет определенный интерес в связи с теорией групп. Кроме того, она приводит нас к идее построения матриц, состоящих не из обычных чисел, а из чисел конечной арифметики.

АГРОНОМИЧЕСКИЕ ИССЛЕДОВАНИЯ[1]

Предположим, что нужно проверить несколько сортов пшеницы. Казалось бы, что может быть легче? Посадите по нескольку зернышек каждого сорта и посмотрите, какой сорт окажется лучшим.

[1] Этот заголовок, так же как и последнюю фразу первого абзаца этой главы, не следует воспринимать буквально: суть дела, конечно, не в «агрономических» обстоятельствах, приведших к этой задаче, а в самом методе статистического контроля.— *Прим. ред.*

Предположим, что самая лучшая пшеница выросла на северо-восточном участке вашего поля. Но, может быть, северо-восточный участок обладает лучшими условиями для произрастания пшеницы, чем любой другой участок: там, скажем, лучше почва; иными словами, урожайность может зависеть вовсе не от сорта пшеницы, а от места, где она насажена. Короче, один урожай — это еще не доказательство.

Способ устранить это возражение очевиден. Один и тот же сорт пшеницы должен быть посажен на всех различных участках, с тем чтобы исключить, насколько это возможно, эффект различия плодородия почв. Научные эксперименты такого рода обычно подвергаются статистическому анализу, в результате которого устанавливается, какой вес следует придавать результатам эксперимента, какова вероятность того, что результат вызван просто внешними, случайными причинами. Следовательно, необходимо выработать некоторый систематический способ смешивания различных сортов пшеницы, чтобы упростить статистический анализ.

Такая система может быть выработана с помощью конечной арифметики. Мы используем арифметику по модулю 5, о которой уже говорилось в начале главы; числами ее являются 0, 1, 2, 3, 4. С помощью этой арифметики мы получим схему посадки на четырех квадратных участках земли пяти сортов пшеницы, урожайность которых нам нужно сравнить.

Располагаем числа следующим образом:

0 1 2 3 4	0 1 2 3 4	0 1 2 3 4	0 1 2 3 4
1	2	3	4
2	4	1	3
3	1	4	2
4	3	2	1

Верхняя строка во всех случаях одинакова. Столбцы взяты из таблицы умножения. При умножении чисел 0, 1, 2, 3, 4 на 1 мы получаем 0, 1, 2, 3, 4. Эти числа и образуют первый столбец. Умножая числа 0, 1, 2, 3, 4 на 2, мы получаем 0, 2, 4, 1, 3, т. е. числа, стоящие во втором столбце. Третий столбец получается путем умножения чисел верхней строки на 3, а последний — на 4.

Затем, следуя правилам сложения, дополняем эти квадраты. Например, во втором квадрате цифра 4 записана в том месте, где

```
0 1 2 3 4
2       .
4       .
1 . . . 4
3
```

должна стоять сумма 1 и 3. Для того чтобы заполнить какую-либо клетку квадрата, мы должны найти числа, стоящие в начале строки и во главе столбца, содержащих эту клетку, сложить их и сумму записать в данную клетку. Под «суммой» мы понимаем, конечно, сумму, образованную по модулю 5, так что, например, 4+3 дает 2.

Таким путем мы получаем четыре квадрата:

0 1 2 3 4	0 1 2 3 4	0 1 2 3 4	0 1 2 3 4
1 2 3 4 0	2 3 4 0 1	3 4 0 1 2	4 0 1 2 3
2 3 4 0 1	4 0 1 2 3	1 2 3 4 0	3 4 0 1 2
3 4 0 1 2	1 2 3 4 0	4 0 1 2 3	2 3 4 0 1
4 0 1 2 3	3 4 0 1 2	2 3 4 0 1	1 2 3 4 0

Это и есть та схема, которой мы должны руководствоваться, рассаживая пшеницу: там, где стоит 1, мы сажаем первый сорт, где 2 — второй и т. д.; там, где встречается цифра 0, мы сажаем пятый сорт.

Всмотритесь в любой квадрат, например в первый, и вы увидите, что цифра 0 встречается в каждой строке, так что если плодородие почвы улучшается в направлении с севера на юг, то это не скажется на средней урожайности сорта 0. По одной цифре 0 содержится также и в каждом столбце, так что постепенное улучшение почвы в направлении с востока на запад также не отразится на результате. То, что справедливо для сорта 0, справедливо и для всех остальных.

Теперь сравним, как расположена цифра 0 в каком-нибудь одном квадрате и в следующем. Вы увидите, что результату опыта не сможет помешать и различие этих расположений. В первом квадрате цифра 0 занимает следующие места

```
* . . . .
. . . . *
. . . * .
. . * . .
. * . . .
```

Взгляните теперь на второй квадрат и найдите, какие числа стоят на этих местах. Это числа 0, 1, 2, 3, 4, причем каждое число встречается один раз.

Все сказанное в равной степени относится к любому числу и к любому квадрату. Возьмите какой-нибудь квадрат и отметьте позиции, которые занимает любое число. Затем перейдите к другому квадрату и посмотрите, какие числа занимают эти же позиции. Вы увидите, что все эти позиции поровну разделены между всеми пятью сортами.

Подобное размещение предметов известно под названием «ортогональных латинских квадратов». Построение таких квадратов кажется тривиальной головоломкой, на самом же деле это задача большой практической важности.

При решении задач прикладной математики часто бывает полезным обратиться к чистой математике и посмотреть, не найдется ли где-нибудь в литературе раздел, в котором уже выведена закономерность, необходимая для решения вашей практической задачи.

РАСШИРЕНИЕ КОНЕЧНОЙ АРИФМЕТИКИ

В арифметике, которой мы только что пользовались, всего пять чисел: 0, 1, 2, 3, 4. Мы можем расширить диапазон чисел следующим образом.

Уравнение $x^2=1$ имеет в нашей арифметике решения 1 и 4, уравнение $x^2=4$ имеет решения 2 и 3, а решением уравнения $x^2=0$ является 0. Уравнения $x^2=2$, $x^2=3$ решений не имеют.

Мы можем ввести новый символ, $\sqrt{2}$, который будет служить решением уравнения $x^2=2$.

Тогда в нашем распоряжении будут следующие 25 величин:

0	1	2	3	4
$\sqrt{2}$	$\sqrt{2}+1$	$\sqrt{2}+2$	$\sqrt{2}+3$	$\sqrt{2}+4$
$2\sqrt{2}$	$2\sqrt{2}+1$	$2\sqrt{2}+2$	$2\sqrt{2}+3$	$2\sqrt{2}+4$
$3\sqrt{2}$	$3\sqrt{2}+1$	$3\sqrt{2}+2$	$3\sqrt{2}+3$	$3\sqrt{3}+4$
$4\sqrt{2}$	$4\sqrt{2}+1$	$4\sqrt{2}+2$	$4\sqrt{2}+3$	$4\sqrt{2}+4$

Эти числа снова образуют конечную арифметику; их можно складывать, вычитать, умножать и делить (за исключением деления на 0), не выходя за пределы этих 25 символов.

Теория таких расширений известна как *теория полей Галуа*. Эта теория есть раздел современной высшей алгебры, что звучит весьма внушительно. Поля Галуа являются, однако, одним из самых простых и легких объектов в современной алгебре, поскольку они состоят из конечного числа элементов[1].

Таблица приведенных выше чисел обладает одним интересным свойством: если умножать число $\sqrt{2}+2$ само на себя, то можно получить последовательно все числа таблицы, кроме 0. Иными словами, любой член таблицы, за исключением нуля, представляет собой степень числа $(\sqrt{2}+2)$ и может быть записан в виде $(\sqrt{2}+2)^n$ с некоторым целым показателем степени n. Этим свойством обладает любое поле[2] Галуа.

---

[1] Не следует, впрочем, думать, что все проблемы, относящиеся к конечным множествам, непременно тривиальны. Вспомним хотя бы человеческий мозг, состоящий из конечного числа клеток — и даже атомов.— *Прим. ред.*

[2] Слово «поле» в алгебре имеет особое значение. Оно не имеет ничего общего с теми полями пшеницы, о которых мы упоминали раньше; оно обозначает некоторое множество символов, которые можно складывать, умножать и делить — то самое, что я называю «арифметикой». [Сказанное, разумеется, не является точным определением понятия поля (см., например, А. Г. Курош. Курс высшей алгебры. М., 1962, стр. 276), но для понимания текста автора таковое и не требуется.— *Ред.*]

Глава четырнадцатая

О ГРУППАХ

Математика двадцать первого века может сильно отличаться от нашей; возможно, школьник начнет изучение алгебры с теории групп подстановок, что он мог бы сделать и сейчас, если бы не установившиеся традиции.

Саймон Ньюкомб, 1893

В конце главы седьмой рассматривались различные группы движений. Доказательства в этой главе, как мне кажется, были очень простыми. Главная трудность при изучении групп заключается не в доказательствах, всегда очень четких и прямых [1], а в том, чтобы уяснить себе цель исследования. В этой главе мы попытаемся, насколько это возможно, рассмотреть вопрос «Как возникла теория групп и зачем она нужна?».

АКСИОМЫ ГРУППЫ

В главе седьмой мы употребляли слово «группа», но не объяснили точно, что оно означает.

Прежде всего, слово «группа» означает совокупность каких-либо объектов (символов, операций или предметов), которые могут комбинироваться некоторым образом друг с другом, причем результатом такой комбинации является снова объект из данной совокупности. Например, если мы умножаем два числа (это тоже способ комбинирования двух чисел), мы ожидаем результат снова в виде числа. Как бы плохо мы ни знали арифметику, нас очень удивило бы, если бы результатом оказался пучок петрушки. Могут быть и комбинации движений: если после поворота на 90° выполнить поворот на 45°, то в результате тело будет повернуто на 135°; суммарный эффект двух движений — это снова движение.

Теория групп лишь постепенно выросла в точную теорию; на первых этапах указанное выше свойство комбинирования было единственным свойством, отличающим группы. Оно часто называлось «групповым свойством».

В настоящее время, для того чтобы некоторая совокупность символов называлась группой, нужно, чтобы она удовлетворяла еще трем требованиям.

1) Совокупность должна содержать символ I, который, сочетаясь с другими символами, не оказывает никакого воздействия на

[1] Такой «педагогический оптимизм» автора не должен, разумеется, приводить читателя к выводу, что в теории групп нет трудных доказательств.— *Прим. ред.*

их, т. е. если X — некоторый символ, то должно быть $IX=X$ и $XI=X$.

Для каждого символа должен существовать ему обратный (какое бы действие ни производил данный символ, обратный должен возвращать все в исходное положение). Например, если в группе содержится операция «умножение на 2», то должна существовать и операция «умножение на $\frac{1}{2}$». Поэтому, например, совокупность операций, заключающихся в умножении только на целые числа, не образует группы.

3) Символы должны подчиняться ассоциативному закону, т. е. если X, Y, Z — некоторые символы группы, то выражения $(XY)Z$ и $X(YZ)$ должны означать одно и то же. В любой группе, элементами которой являются операции, этот закон выполняется автоматически. Предположим, например, что этими операциями являются операции перемещения. Пусть Z (рис. 83) перемещает объект из положения (1) в положение (2), Y — из положения (2) в положение (3), а X — из (3) в (4). XY означает последовательное выполнение сначала операции Y, а затем операции X (вспомните, что порядок записи символов обратен порядку выполнения операции), т. е. операция XY перемещает предмет из положения (2) в положение (4). $(XY)Z$ означает, что сначала выполняется операция Z, а затем операция XY. Поскольку Z перемещает предмет из положения (1) в положение (2), а XY — из (2) в (4), то суммарный эффект выражается в перемещении предмета из (1) в (4). Аналогично вы можете убедиться, что операция $X(YZ)$ тоже производит перемещение предмета из (1) в (4).

Итак, все сводится к следующему: сказать «от (1) к (2)» и после этого, не *переводя дыхания*, «и от (2) к (3), а затем к (4)» — все равно, что сначала произнести на одном дыхании «от (1) к (2), а затем к (3)», а потом спокойно «и от (3) к (4)».

Первое утверждение соответствует $(XY)Z$, а второе — $X(YZ)$, как это явствует из рис. 83.

Так как мы будем заниматься главным образом группами операций, вы можете не обращать внимания на требование 3); оно выполняется для таких групп автоматически.

Итак, группа операций должна обладать следующими свойствами: последовательное выполнение двух операций группы равносильно выполнению некоторой одной операции из той же группы; имеется операция I, которая оставляет объекты без изменения; какую бы операцию из группы вы ни взяли, обратная к ней операция снова принадлежит группе.

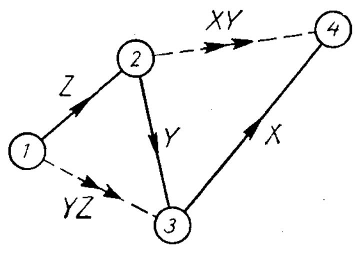


Рис. 83.

Самым замечательным является то, что на основе этих нескольких предположений можно построить грандиозную теорию.

Возможно, вы захотите вернуться к главе седьмой и убедиться, что все эти требования выполняются в случае группы прямоугольника и группы равностороннего треугольника. Какая операция в группе прямоугольника является обратной для p? Это — сама p.

ПРОИСХОЖДЕНИЕ ТЕОРИИ ГРУПП

Необходимость в теории групп возникла в связи с вопросом, который сам по себе не имеет никакой практической важности, но как образец метода исследования играет исключительную роль для всех областей математики.

Это была задача решения алгебраических уравнений. Линейные и квадратные уравнения были решены еще в древние века. Уравнения третьей и четвертой степеней были решены незадолго до 1550 года. И на этом дело застряло. Много математиков пыталось решить уравнение пятой степени, но безуспешно. Была достигнута некоторая унификация, было показано, что все методы, которыми решаются уравнения первой, второй, третьей и четвертой степеней, следует рассматривать как частные случаи одного-единственного метода, но *этот метод оказался непригодным для решения уравнений пятой степени.*

Решить задачу — значит свести ее к более простой задаче. Постепенно люди начали приходить к выводу, что, возможно, уравнение пятой степени нельзя свести к чему-либо более простому. И, наконец, было окончательно доказано, что так оно и есть.

ЗАДАЧИ-АТОМЫ

Таким образом, уравнение пятой степени — это своего рода математический атом, нечто такое, что нельзя разделить на более простые вещи алгебраическим путем.

Для иллюстрации понятия «задачи-атомы» можно взять совсем простой пример. Если мы хотим решить уравнение $x^6=2$, мы можем это сделать в два шага.

По таблице квадратных корней мы можем найти $\sqrt{2} = 1{,}4142$, а затем по таблице кубических корней $\sqrt[3]{1{,}4142} = 1{,}1223$.

Таблица корней шестой степени становится излишней; корень шестой степени может быть найден с помощью таблиц квадратных и кубических корней.

Алгебраически это означает, что решение уравнения $x^6=2$ может быть заменено решением уравнений

$$y^2 = 2, \qquad x^3 = y.$$

Следовательно, решение уравнения $x^6=2$ не представляет собой задачу-атом, его можно разбить на две задачи.

С другой стороны, с уравнением $x^2=2$ мы не можем поступить таким же образом. Это задача-атом. Ее нельзя свести к двум более простым задачам.

Поэтому, приступая к решению любого алгебраического уравнения, нужно задаться вопросом, можно ли это уравнение свести к двум более простым уравнениям. Если нет, то это (по крайней мере, с точки зрения алгебры) задача-атом, и напрасной тратой времени будет выискивать способы ее решения. Стоит заметить, что математики бились более 250 лет над проблемой решения уравнения пятой степени, прежде чем они поняли, что пытались сделать невозможное.

Задача-атом не является абсолютно неразрешимой. Если ввести в дело тяжелую артиллерию, то она перестает быть таковой. Например, квадратный корень можно найти с помощью логарифмов. Задача только относительно неразрешима, неразрешима с помощью данных инструментов — вроде деревянного плуга из пятой главы. Большая экономия времени достигается благодаря теории, которая показывает невозможность решения задачи данным методом. Тогда сразу начинают отыскивать новый метод, вместо того чтобы тратить века на бесконечное варьирование в пределах старого метода.

Галуа, блестящий математик, погибший в возрасте 21 года (в 1832 г.), показал, что с любым алгебраическим уравнением связана некоторая группа, и, исследуя эту группу, можно сказать, является ли данное уравнение задачей-атомом или же нет. Если группа, как говорят, *непростая,* то уравнение может быть сведено к двум более простым уравнениям; если же группа *простая,* то уравнение упростить нельзя (задача-атом).

Овладеть теорией Галуа не очень просто. В работах по современной алгебре эта теория представлена в форме, несколько отличной от той, которую создал сам Галуа.

Здесь я могу дать лишь очень слабое представление об этой теории, описав с помощью одного-двух примеров, как уравнение может быть связано с группой и что такое простая и непростая группы. Возможно, эти примеры покажутся вам своего рода трюкачеством, но у меня нет места, чтобы изложить здесь теорию, которая привела меня именно к этим примерам.

ЦИКЛИЧЕСКОЕ УРАВНЕНИЕ

Прежде всего рассмотрим уравнение $x^3-3x+1=0$, корни которого мы обозначим буквами a, b, c. Собственно говоря, эти корни легко записать в тригонометрическом виде:

$$a=2\cos 40°,\ b=2\cos 80° \text{ и } c=2\cos 160°.$$

Посмотрите на значения углов и вы увидите, что 80° — это дважды 40°, а 160° — это дважды 80°. Интересно посмотреть, что получится,

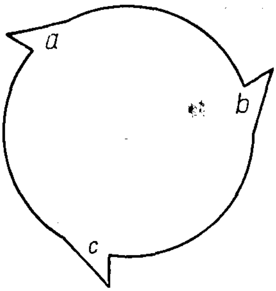


Рис. 84.

если мы еще раз удвоим значение угла. Дважды 160° — это 320°, а косинус 320° равен косинусу 40°, так что здесь у нас получилось нечто вроде змеи, кусающей собственный хвост. Эта операция (удвоение величины угла) ведет нас от a к b, затем от b к c и от c к a. Буквы a, b, c обладают симметрией такого же рода, как изображенное здесь трезубчатое колесо (рис. 84).

Мы можем повернуть это колесо таким образом, чтобы на месте, где раньше находилось a, оказалось b, на месте b оказалось c, а на месте c оказалось a.

Соотношения между a, b и c можно выразить с помощью равенств

(1) $$b = a^2 - 2,$$

(2) $$c = b^2 - 2,$$

(3) $$a = c^2 - 2.$$

Бросается в глаза симметрия этих соотношений; действительно, ни одна буква не может пожаловаться на несправедливое отношение. И все-таки полной симметрии нет. Например, в равенстве (1) мы не можем поменять местами буквы a и b, так как равенство потеряет силу. В действительности выражение a через b дается равенством

(4) $$a = -b^2 - b + 2,$$

к которому, разумеется, можно добавить

(5) $$b = -c^2 - c + 2,$$

(6) $$c = -a^2 - a + 2.$$

Это можно сравнить с ситуацией, которая иногда встречается в играх: B может победить A, C может победить B, A может победить C[1]. Ситуация симметрична, но эта симметрия, если можно так выразиться, односторонняя. Утверждения «A проигрывает B», «B проигрывает C», «C проигрывает A» обладают такого же рода симметрией.

Эта ситуация обладает полной симметрией в случае, когда B играет вничью с A, C с B и A с C. Тогда можно поменять местами

[1] Один африканский школьник однажды спросил меня, каким образом, имея в виду возможность такой ситуации, можно утверждать правильность высказывания: «Если x больше y и y больше z, то x больше z». Как и в большинстве случаев, когда вопрос задает ребенок, я обнаружил, что мне пришлось немало подумать, прежде чем я смог ответить на этот вопрос.

A и B, и утверждение «B играет вничью с A» столь же справедливо, как и утверждение «A играет вничью с B».

Полной симметрией обладает группа равностороннего треугольника, которую мы рассматривали в главе седьмой.

Тот тип симметрии, которым обладают зубчатое колесо, уравнения (1), (2), (3) и игра, когда A проигрывает B, B проигрывает C и C проигрывает A, называется *циклической симметрией.*

Зубчатое колесо можно повернуть вокруг его оси, но нельзя вынуть из плоскости и положить на другую сторону, как это можно сделать с равносторонним треугольником.

ЦИКЛИЧЕСКАЯ ГРУППА

Когда мы встречаемся с этим типом симметрии, напоминающим нам собаку, кусающую хвост, ясно, что если справедливо некоторое утверждение об игроках A, B, C, то справедливо такое же утверждение об игроках B, C, A. Аналогично, если из уравнений (1), (2), (3) мы можем заключить, что $f(a, b, c)=0$, то мы сможем также вывести из них $f(b, c, a)=0$.

С помощью какой операции второе равенство получается из первого? Где бы ни встретилось a в первом равенстве, мы его стираем и пишем b, везде на месте b мы пишем c, а на месте c — a.

Эту операцию часто обозначают символом (abc). Каждая буква уступает свое место следующей за ней букве внутри скобок: a уступает место b, b уступает c, c — это последняя буква в скобке, и, строго говоря, за ней ничего не следует, но мы, однако, будем считать, что первая буква в скобке следует за последней, поэтому c уступает свое место a.

Для краткости я буду обозначать операцию (abc) буквой ω. Тогда для любой функции $f(a, b, c)$ имеем: $\omega f(a, b, c)=f(b, c, a)$. Это следует читать: «Операция ω, воздействуя на $f(a, b, c)$, дает $f(b, c, a)$».

Эту операцию можно, конечно, выполнить дважды. Применяя ее еще раз, получим $\omega^2 f(a, b, c)=f(c, a, b)$. Выполнив операцию три раза, получаем $\omega^3 f(a. b, c)=f(a, b, c)$, т. е. троекратное выполнение операции ω возвращает нас в исходное положение. Мы можем поэтому написать $\omega^3=I$.

Операции I, ω, ω^2 образуют группу, известную под названием «циклической группы третьего порядка», или, короче, C_3. Эта группа связана с уравнением $x^3-3x+1=0$ и является группой Галуа для этого уравнения.

Если между корнями a, b, c указанного выше уравнения имеется некоторое соотношение и мы применяем к этому соотношению любую операцию группы Галуа, то в качестве результата мы получим тоже верное соотношение.

Например, приведенное ранее в этом разделе равенство (4) является верным соотношением между a, b и c. Применив операцию I, мы оставим равенство (4) без изменения. Если же мы применим к нему операцию ω, то получим равенство (5), которое тоже является верным соотношением между корнями. Применение операции ω^2 приводит нас к соотношению (6).

Я уверен, что, когда вы читали раздел этой главы, где говорилось о соотношении (4), вы сразу почувствовали, что у него должны быть сопутствующие соотношения (5) и (6), и если бы я не привел их там, то вы могли бы написать их сами. Поэтому не расценивайте мое объяснение операции ω и ω^2 как нечто совершенно новое: операции ω и ω^2 — *это просто то, что вы делали, когда увидели, что уравнению* (4) *должны сопутствовать уравнения* (5) *и* (6).

Введение операций ω и ω^2 имеет целью заменить смутное ощущение закономерности строгой и сознательной идеей, которая может быть выражена с помощью символов и может служить основой для математической теории.

Группа Галуа состоит из перестановок, которые можно произвести над корнями уравнения, причем так, что, если некоторое утверждение было верно до перестановки, оно будет верно и после нее.

ОГРАНИЧЕНИЕ КРУГА УТВЕРЖДЕНИЙ

Мы должны с известной осторожностью употреблять слово «утверждение». Мы уже говорили, что при известных обстоятельствах (рассмотренных ранее) утверждение, сделанное относительно трех игроков A, B, C, справедливо и для B, C, A. Но ясно, что это не относится к утверждению типа «у A красный нос». Отсюда мы не можем заключить, что у B и у C носы тоже красные. Утверждение должно относиться к игре. Однако и не всякое относящееся к игре утверждение справедливо для B и C, если оно справедливо для A. Например, из утверждения «у A сильный удар справа» вовсе не следует, что у B, играющего сильнее A, тоже сильный удар справа. Возможно, B побеждает благодаря своей ловкости, не позволяя A применить свой сильный удар справа. В действительности круг утверждений должен ограничиваться только теми утверждениями, которые касаются результатов игры. Если мы начнем вдаваться в подробности, выяснять, каким методом был достигнут результат, то все эти утверждения для наших теперешних целей непригодны.

Подобного же рода ограничения мы должны наложить и на наши алгебраические задачи. Рассмотрим, например, утверждение, что корень a равен $2 \cos 40°$, т. е. $= 1{,}53208\ldots$. Если к равенству $a = 1{,}53208 \ldots$ мы применим операцию ω, то, заменив a на b, получим $b = 1{,}53208$, *что неверно*. Мы должны ограничить круг наших функций $f(a, b, c)$ многочленами от abc, в которых коэффициентами могут быть *только целые числа*. Так, соотношение вида $ac - abc + a^2 + a - 2 = 0$ имеет право называться «утверждением»; соотношения, куда входят выражения вроде $a+b\sqrt{2}$ или $c-\pi$, содержащие иррациональные числа $\sqrt{2}$ и π, нам придется исключить. Дроби, однако, не мешают. Например, $\frac{1}{2}(a^2+a+c)-1=0$ означает то же самое, что $a^2+a+c-2=0$, и имеет право называться «утверждением». В действительности это уравнение (6).

УТВЕРЖДЕНИЯ ОТНОСИТЕЛЬНО $\sqrt{2}$

Употребляя слово «утверждение» все в том же смысле, давайте посмотрим, какое утверждение может быть сделано относительно $\sqrt{2}$. Это утверждение может быть записано в виде $f(\sqrt{2})=0$ или, если угодно, $f(x)=0$, где $x^2=2$, а функция $f(x)$ есть многочлен с целыми коэффициентами. Предположим, мы делим $f(x)$ на x^2-2. Получится некоторое частное, скажем $\varphi(x)$, и линейный остаток, скажем $px+q$, где p и q — целые числа; что это на самом деле так, вы можете проверить, взяв любой многочлен с целыми коэффициентами, скажем

$$3x^4-7x^3+4x^2+9x+5,$$

и разделив его на x^2-2. Нигде в процессе деления дроби не возникнут. Когда мы делим 23 на 7, то мы получаем в частном 3 и в остатке 2; все это можно записать в виде одного равенства: $23=7\times 3+2$. Подобным же образом интересующее нас деление может быть представлено равенством

$$f(x)=(x^2-2)\varphi(x)+px+q.$$

В этом равенстве мы можем придать символу x любое значение. Положим, $x=\sqrt{2}$. Тогда x^2-2 станет равным нулю. Поскольку $f(\sqrt{2})=0$, то $f(x)$ тоже станет равно 0. Наше равенство сведется к

$$0=p\sqrt{2}+q,$$

откуда, казалось бы, $\sqrt{2}=-\frac{q}{b}$. Но ведь хорошо известно, что $\sqrt{2}$ не может быть представлен рациональной дробью; получается как будто противоречие. Единственную лазейку представляет нам возможность $p=0$, $q=0$. Поскольку это единственный путь избежать абсурда, то так оно и должно быть. Следовательно, $p=0$, $q=0$, т. е. весь остаток $px+q$ равен нулю. Это означает, что многочлен $f(x)$ должен делиться без остатка на x^2-2, т. е. $f(x)=(x^2-2)\varphi(x)$.

Итак, мы выяснили, что утверждение $f(\sqrt{2})=0$ может быть сделано только в том случае, если $f(x)$ имеет вид $(x^2-2)\varphi(x)$.

А если это так, то, положив $x=-\sqrt{2}$, мы получим $f(-\sqrt{2})=0$. Следовательно, *если какое-либо утверждение может быть сделано* *относительно* $\sqrt{2}$, *то такое же утверждение можно сделать и относительно* $-\sqrt{2}$. Термин «утверждение» мы здесь, конечно, употребляем в том ограниченном смысле, как это объяснено раньше; сформулированный выше принцип превратился бы в бессмыслицу, если мы применили бы его к предложению типа «$\sqrt{2}$ — положительное число». При правильном использовании этого принципа мы приходим к результатам вроде следующего: если $x = \sqrt{2}$ удовлетворяет уравнению $x^3 - x^2 - 2x + 2 = 0$, то $x = -\sqrt{2}$ тоже удовлетворяет этому уравнению, ибо здесь мы имеем *как раз тот тип уравнения* («утверждение»), к которому применим наш принцип.

УРАВНЕНИЕ $X^4 - 10X^2 + 1 = 0$

Принцип, который мы применили выше к $\sqrt{2}$, может быть распространен; благодаря такому распространению можно, например, найти группу Галуа для уравнения $x^4 - 10x^2 + 1 = 0$, имеющего корни $a = \sqrt{2} + \sqrt{3}$; $b = \sqrt{2} - \sqrt{3}$; $c = -\sqrt{2} + \sqrt{3}$; $d = -\sqrt{2} - \sqrt{3}$. Я не собираюсь входить в подробности доказательства, но если в любое утверждение относительно чисел a, b, c, d подставить вместо $\sqrt{2}$ число $-\sqrt{2}$, утверждение останется верным. Эффект замены числа $\sqrt{2}$ на $-\sqrt{2}$ заключается в том, что значение $a = \sqrt{2} + \sqrt{3}$ изменяется на $-\sqrt{2} + \sqrt{3}$, т. е. a превращается в c. Аналогично c превращается в b, b превращается в d, а d — в a.

При замене $\sqrt{2}$ на $-\sqrt{2}$ меняются местами a с c и b с d, так что $f(a, b, c, d)$ переходит в $f(c, d, a, b)$.

Далее, нет такого утверждения, которое, будучи верно относительно $\sqrt{3}$, не было бы верно и относительно $-\sqrt{3}$. Мы можем поэтому заменять $\sqrt{3}$ на $-\sqrt{3}$. Это приведет к тому, что a поменяется местом с b, а c—с d; $f(a, b, c, d)$ перейдет в $f(b, a, d, c)$.

И, наконец, можно произвести обе замены одновременно. Мы можем заменить $\sqrt{2}$ на $-\sqrt{2}$, а $\sqrt{3}$ на $-\sqrt{3}$. Это превратит a в d, b в c, c в b и d в a. В результате $f(a, b, c, d)$ перейдет в $f(d, c, b, a)$.

Все перечисленные выше операции, совместно с операцией I, оставляющей буквы a, b, c, d на своих местах, составляют группу Галуа для этого уравнения. Если $f(a, c, b, d) = 0$ есть правильное утверждение относительно корней этого уравнения, то $f(c, d, a, b) = 0$, $f(b, a, d, c) = 0$ и $f(d, c, b, a) = 0$ будут тоже правильными утверждениями.

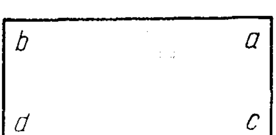


Рис. 85.

С этой группой мы уже встречались раньше. Если написать буквы a, b, c, d в углах прямоугольника на рис. 85, то перестановки букв, найденные выше для уравнения $x^4-10x^2+1=0$, в точности совпадают с теми, которые возникают, когда прямоугольник вынимается из рамки и размещается в ней другим образом. Группа прямоугольника была первой группой, которую мы рассматривали в главе седьмой.

Однако как в случае уравнения $x^4-10x^2+1=0$, так и в случае уравнения $x^3-3x+1=0$, рассмотренного ранее, корни уравнения нам были известны до того, как мы составили группу Галуа. Но, оказывается, возможно (с помощью довольно сложной процедуры) определить группу Галуа, не решая самого уравнения. Если бы этого нельзя было сделать, то теория оказалась бы бесполезной, потому что все значение групп Галуа заключается в том, что они сообщают нам, *насколько трудным является путь к решению уравнения и как мы должны действовать для этой цели.*

ПРОСТЫЕ И НЕПРОСТЫЕ ГРУППЫ

Каким же образом группа уравнения сообщает нам информацию о его решении? Я отвечу на этот вопрос только в одном смысле. Я покажу, как группа дает нам возможность определить, является ли данная задача задачей-атомом или же нет, т. е. можно ли ее свести к решению более простых задач. Я постараюсь объяснить, что́ мы ищем в группе, не пытаясь обосновать свою мысль.

Рассмотрим уравнение, для которого группой Галуа является группа равностороннего треугольника.

Уравнение, которому принадлежит эта группа, — это уравнение $x^3=2$. Возможно, с первого взгляда вы решите, что эта задача-атом; кажется, что не существует способов сведения ее к более простой. Но вы должны помнить, что кубическое уравнение имеет три корня. Полное решение уравнения дается тремя числами a, b, c, где

$$a=\sqrt[3]{2}\cdot$$

$$b=\sqrt[3]{2}\left\{\frac{-1+\sqrt{-3}}{2}\right\},$$

$$c=\sqrt[3]{2}\left\{\frac{-1-\sqrt{-3}}{2}\right\}.$$

Таким образом, полное решение уравнения включает нахождение двух совершенно различных вещей, а именно $\sqrt[3]{2}$ и $\sqrt{-3}$. Значит, задача разбивается на две более простые задачи и не является задачей-атомом.

Как эта способность распадаться на более простые задачи проявляется в группе Галуа данного уравнения? Вспомним, что эта группа является группой равностороннего треугольника и представляется таблицей:

	I	ω	ω^2	p	q	r
I	I	ω	ω^2	p	q	r
ω	ω	ω^2	I	r	p	q
ω^2	ω^2	I	ω	q	r	p
p	p	q	r	I	ω	ω^2
q	q	r	p	ω^2	I	ω
r	r	p	q	ω	ω^2	I

Взглянув на таблицу, мы подмечаем ее характерную особенность. Она распадается на квадраты. Буквы p, q, r встречаются в северо-восточном и юго-западном углах, а символы I, ω, ω^2 — на северо-западе и юго-востоке. Если бы мы написали буквы I, ω, ω^2 красными чернилами, а буквы p, q, r — черными, то мы бы особенно наглядно увидели, что буквы размещены квадратами. Если смотреть на таблицу с такого расстояния, что видны были бы только *цвета*, то таблица нам показалась бы такой:

	Красный	Черный
Красный	Красный	Черный
Черный	Черный	Красный

Эта табличка сама по себе является моделью группы, но группы меньшей, чем исходная группа. Мы можем прочесть ее как «красный×красный-красный, красный×черный-черный» и т. д. Эта маленькая группа имеет то же строение, что и группа, приведенная в последнем разделе главы седьмой,

	I	k
I	I	k
k	k	I

или как таблица умножения для чисел 1 и —1:

	1	−1
1	1	−1
−1	−1	1

или же как таблица сложения для понятий «чет» и «нечет»:

	чет	нечет
чет	чет	нечет
нечет	нечет	чет

Поэтому большую группу, состоящую из шести элементов, можно рассматривать как дальнейшее развитие модели для малой группы. Мы можем даже сказать, что большая группа *отображена* на меньшую путем следующего соответствия:

$$\underbrace{I\ \omega\ \omega^2}_{\text{красный}} \qquad\qquad \underbrace{p\ q\ r}_{\text{черный}}$$

Тот факт, что уравнение может быть разбито на более простые уравнения, проявляется в структуре таблицы группы; она обязательно содержит структуру меньшей группы.

В начале этой главы мы рассматривали уравнение

$$x^4 - 10x^2 + 1 = 0,$$

которое, будучи разрешимым с помощью величин $\sqrt{2}$ и $\sqrt{3}$, не является задачей-атомом. Группа Галуа для этого уравнения — это группа прямоугольника, имеющая таблицу:

	I	p	q	r
I	I	p	q	r
p	p	I	r	q
q	q	r	I	p
r	r	q	p	I

Тот факт, что уравнение не является задачей-атомом, немедленно усматривается при написании букв I и p красным, а q и r — черным цветом.

Когда структура группы представляет в этом смысле модель меньшей группы, то группа называется *непростой.* Если же группа таким свойством не обладает, она называется *простой.* Простая группа соответствует задаче-атому.

Группа I, ω, ω^2, с которой мы встретились в связи с циклическим уравнением $x^3 - 3x + 1 = 0$, является простой. Она содержит 3 элемента, а 3 — это простое число. Любая группа, содержащая простое число элементов, неизбежно является простой. Простые группы, количество элементов которых выражается непростым числом, интересны и довольно редки.

Группа, известная под названием группы икосаэдра и содержащая 60 элементов, является самой маленькой группой подобного типа. Неразрешимость уравнения пятой степени тесно связана с этой группой.

Далее идет группа, содержащая 168 элементов, которая связана с «конечной геометрией», описанной в главе тринадцатой.

АНАЛОГИ ТЕОРИИ ГАЛУА

Теория Галуа интересна. Но полезна ли она? Ответ на этот вопрос служит прекрасной иллюстрацией того, какого рода полезность может иметь математическая теория. Теория Галуа непосредственно на практике нигде не применяется. Никому на практике не нужно решать уравнения пятой степени. Если бы такое уравнение возникло в практической работе, вы начертили бы график и посмотрели бы, в каких точках эта кривая пересекается с осью x.

Истинная ценность теории Галуа заключается в том, что она служит моделью для почти любого рода исследований. Раньше математик спрашивал: «Как найти способ решить эту задачу?» И если он не мог найти его сегодня, он продолжал свои поиски завтра. Но с появлением теории Галуа изменилась сама постановка задачи. Мы больше не гадаем, существует ли вообще способ решения. Мы спрашиваем: «Есть ли основания предположить, что задача может быть решена такими-то средствами? Можно ли разбить ее на более простые задачи? Что делает задачу разрешимой, и как мы можем проверить эту разрешимость?». Мы не пытаемся более изобретать; мы пробуем познать природу задачи, с которой нам приходится иметь дело.

В отличие от алгебраических уравнений решение дифференциальных уравнений — дело большой практической важности. Между 1883 и 1892 гг. Пикар и Вессио создали теорию дифференциальных уравнений, очень напоминающую теорию Галуа для алгебраических уравнений. Похоже на то, что она возникла под влиянием теории Галуа.

Очевидной целью математического исследования является распространение идей Галуа на все типы математических задач: нужно показать, какие из них являются задачами-атомами и как узнавать составные задачи.

Любая подобная теория представляет ценность в основном для математика-теоретика. Она сберегает ему время, которое он потратил бы на поиски решения задачи с помощью совершенно непригодных в данном случае методов. Она предполагает некоторый систематический подход к решению новых задач. У инженера редко находится время для фундаментальных математических исследований, он обычно довольствуется методами, развитыми математиками, применяя их для нужд практических вычислений. Теория Галуа не имеет непосредственного применения в инженерной практике.

ОБ АВТОРЕ

У. У. Сойер родился в 1911 году. Он получил образование в Хайпетской школе и в Сент-Джонс-Колледже в Кембридже, где занимался главным образом математическими аспектами квантовой теории и теории относительности. Читал лекции по математике в Данди и Манчестере, а в 1945—1947 гг. был деканом математического факультета Лейстерского технологического колледжа. В это время он занимался исследованиями по применению математики в промышленности и поисками новых путей преподавания математики студентам технических учебных заведений. Опыт его работы частично описан в книгах «Математика в теории и практике» и «Проект и работа».

В 1948—1950 гг. У. У. Сойер — декан математического факультета университетского колледжа Золотого Берега. В настоящее время [1] читает лекции по математике в Кентерберийском колледже в Кристчерче, в Новой Зеландии.

Пытаясь бороться с нехваткой преподавателей математики в Новой Зеландии, он организовал добровольные математические общества школьников. Надеется, что интерес, вызванный этими обществами, заставит его учеников выбрать путь учителя математики. Первая книга Сойера «Восторг математика» была опубликована в 1943 г.

---

[1] Данная аннотация к английскому изданию относится к 1955 г.— *Прим. ред.*

СОДЕРЖАНИЕ

У. У. Сойер

ПРЕЛЮДИЯ К МАТЕМАТИКЕ

Редактор *Ю. А. Гастев*. Обложка *Н. Н. Румянцева*.
Художественный редактор *Е. Н. Карасик*.
Технический редактор *Е. К. Полукарова*.
Корректор *Т. М. Графовская*.

Сдано в набор 16/IV 1971 г. Подписано к печати 20/I 1972 г. 60×90[1]/16. Печ. л. 12,0. Уч.-изд. л. 12,22. Тираж 80 000 экз. Бум. тип. № 2. (План 1972 г. № 120). Цена 57 коп.

Издательство «Просвещение» Комитета по печати при Совете Министров РСФСР. Москва, 3-й проезд Марьиной рощи, 41.

Ордена Трудового Красного Знамени типография издательства ЦК КП Белоруссии, Минск, Ленинский проспект, 79. Заказ 228.