HANS FREUDENTHAL

MATHEMATIK IN WISSENSCHAFT UND ALLTAG

KINDLER VERLAG GMBH MÜNCHEN, 1968

ГАНС ФРЕЙДЕНТАЛЬ

МАТЕМАТИКА В НАУКЕ И ВОКРУГ НАС

Перевод с немецкого
Ю. А. Данилова

ИЗДАТЕЛЬСТВО «МИР» МОСКВА 1977

51
Ф 86

Фрейденталь Г.

Ф 56 Математика в науке и вокруг нас. Пер. с нем. Ю. А. Данилова. М., «Мир», 1977.

261 с. с ил. (В мире науки и техники)

Математика издавна служила людям надежным подспорьем в коммерческих расчетах, помогала навигаторам определять положение судна в море, землемерам — измерять земельные участки, астрономам — составлять календари.

О развитии математики, о том, какое место занимает она в современной науке и жизни каждого из нас, со знанием дела, живо и увлекательно рассказывает известный голландский математик и педагог Ганс Фрейденталь. Его книгу с большим интересом и пользой прочтут люди самого разного возраста, профессий и увлечений.

Ф $\frac{20201-176}{041(01)-77}$ 176—77 51

Редакция научно-популярной и научно-фантастической литературы

Предисловие переводчика

«Введение к этой работе», которым открывается замечательное сочинение Иоганна Кеплера «Новая астрономия, или физика небес» (отчет об открытии двух первых законов движения планет, носящих ныне имя великого преобразователя астрономии и известных каждому школьнику), содержит поистине пророческое признание: «Тяжкий жребий — писать в наши дни математические книги... Если не соблюдать надлежащей строгости в формулировках теорем, пояснениях, доказательствах и следствиях, то книгу нельзя считать математической. Если неукоснительно соблюдать все требования строгости, то чтение книги становится весьма затруднительным...»

С дилеммой «строгость или доступность» сталкивается каждый, кто пишет любую книгу по математике, в особенности книгу популярную. Автор книги, предлагаемой вниманию читателя, разрешил эту дилемму весьма искусно. Не ставя перед собой неразрешимой задачи дать полную картину многообразной жизни современной математики, он мастерски, семью широкими мазками набрасывает портрет древней, но вечной юной науки, и в наш век всеобщей математизации остающейся для большинства людей прекрасной незнакомкой. Умело соразмеряя строгость и описательный подход, Ганс Фрейденталь создает картину, передающую характер «натуры» точнее, чем унылое перечисление имен, фактов, методов и результатов, подобно тому как портрет кисти выдающегося мастера передает своеобразие изображенного на нем лица более глубоко и верно, чем портрет того же лица, созданный фотороботом. И хотя выбор тем для семи глав книги продиктован личными вкусами и научными привязанностями автора, картина современной математики предстает перед читателем достаточно полно Нет необходимости останавливаться на содержани: книги более подробно — оно ярко и точно охаракте ризовано в предисловии самого автора. Поэтому м ограничимся лишь тем, что представим его читател:

Выдающийся голландский математик Ганс Фре. денталь родился в 1905 г. в Германии. Математиче ское образование получил в Берлинском университете где проучился с 1923 по 1930 г. (1927 г. Г. Фрейденталі провел в Сорбонне). Докторскую диссертацию защи тил в 1930 г. До 1946 г. был ассистентом Амстердам ского университета, а с 1946 г. занял пост профессор: по кафедре чистой и прикладной математики Утрехт ского университета. Член Королевской Нидерландско академии, иностранный член Королевской Фламанд ской академии, почетный доктор Берлинского универ ситета. В математике Г. Фрейденталь следует «роман тическому направлению». Его научному творчеств свойственна широта и разнообразие тем, богатство идей и методов. Автор более чем 200 работ по тополс гии, теории групп Ли, анализу, геометрии, математи ческой логике, философии и истории наук (в том чис ле книги «Линкос: проект языка для общения с ино планетными цивилизациями»), Ганс Фрейденталі известен также как блестящий популяризатор и орга низатор ряда международных симпозиумов.

Ю. Данилов

Введение

Люди умели считать и занимались изучением геометрических фигур еще в те времена, когда они не научились писать. С возникновением письменности появились числа и весьма скоро вслед за ними — высокоразвитая математика. За три тысячелетия до новой эры вавилоняне умели решать квадратные уравнения и знали теорему, которую ныне совершенно необоснованно мы называем теоремой Пифагора. Для чего была нужна математика в древности? Она являлась надежным подспорьем при коммерческих расчетах, помогала навигаторам определять положение судна в море, землемерам — измерять земельные участки, астрономам — составлять календари и служила многим другим столь же «земным» целям. Но с самого начала математика вышла далеко за пределы практических потребностей человечества. Игра с числами и фигурами была не только средством, но и самоцелью.

По мере развития человечества появлялись все новые и новые области приложения математики. Долгое время важнейшей из них считалась астрономия. Возникшая в первом тысячелетии до новой эры в Вавилоне и перенятая затем греками, она обрела форму, которая сохранялась неизменной более тысячелетия.

Чистая математика всегда далеко опережала развитие своих приложений как по содержанию, так и по форме. Греки не только развили и дополнили наследие вавилонян, но и создали нечто совершенно новое. Они превратили математику в логическую систему, в которой от исходных предположений с помощью логического вывода, называемого доказательством, можно переходить к заключениям. Взгляд на математику как на логическую систему в значительной мере сохранился и поныне. В математике ничто не требуется принимать на веру. Математик ничего не рассказывает нам о реальности. Он говорит лишь о том, что одно следует из другого, и показывает, как именно это происходит. В истинности логических построений математика каждый может убедиться своим собственным разумом.

Постепенно математику все шире стали использовать и другие науки. Первой наукой, глубоко постигшей преимущества математического подхода, была астрономия. Небесная механика нуждалась в действенных математических методах, которые отчасти и были разработаны для удовлетворения ее запросов.

Для применения математики открывалось все больше самых разнообразных возможностей. И все же еще в начале XX в. из сотни выпускников гимназии вряд ли можно было встретить одного, кому хотя бы раз пришлось применить ту математику, которую он изучал на школьной скамье.

Для чего же тогда изучали математику? Утверждают, что математика дисциплинирует и развивает ум. Это явное преувеличение, но оно содержит зерно истины. Человек, искушенный в математике, как правило, даже не сознавая, использует методы математического мышления на каждом шагу, по любому поводу.

В чем, собственно, состоят методы математического мышления? На этот вопрос, так же как и на вопрос: «Как люди плавают?», невозможно дать убедительный ответ. Впрочем, всякому, кто пожелает научиться плавать, можно показать, как плавают другие, чтобы он мог подражать их движениям. Точно так же человек может научиться математически мыслить, подражая тем, кто уже овладел этим искусством. Тому, как мыслят математики, и посвящена эта книга.

Что же такое математика, на которой мы собираемся сосредоточить все свое внимание? Это также можно показать лишь на примерах. Однако о многих разновидностях человеческой деятельности сразу можно сказать: «Это не математика». Таковы, например, вычисления. Правда, было и такое время, когда ими владели лишь ученые, но оно давно прошло. Еще задолго до появления вычислительных машин численные расчеты были организованы так, чтобы их можно было производить чисто механически, не задумываясь. Жонглирование алгебраическими формулами тоже не является математикой, хотя умение свободно обращаться с формулами, так же как и умение производить численные выкладки, может оказаться небесполезным и для математика. Создание численных методов и нашей алгебры, разумеется, было делом математиков. Они и поныне неустанно трудятся над разработкой простых и надежных методов расчета, называемых алгоритмами, для самых разнообразных областей математики. Но коль скоро алгоритм создан, все остальное можно поручить вычислительной машине. То, чем занимается вычислительная машина, не является математикой, но для того, чтобы вычислительная машина заработала, нужны математика и математики.

И все же математик должен изучить даже ту чисто техническую сторону своей науки, которую он на практике поручил бы вычислительной машине, подобно тому как в школе нам приходится учиться писать обыкновенной ручкой, хотя быстрее печатать на пишущей машинке. Всякий, кто пожелает освоить техническую сторону математики, должен обратиться к учебнику: там все изложено систематически. Мы предлагаем читателю не учебник. Не становясь математиком, читатель совершит путешествие по нескольким разделам этой науки, подобно тому как можно побывать с экскурсией на верфи, в газетном издательстве или в школе, обойти там все помещения от конструкторского бюро до стапеля, от наборного цеха до экспедиции, заглянуть во все классы, не став при этом ни судостроителем, ни издателем, ни учителем. Вы подойдете к столу математика, увидите математика за работой, загляните ему через плечо и, когда сочтете, что увидели достаточно, пойдете дальше, в другую комнату, где сидит другой математик, которому также можно заглянуть через плечо. Гораздо сложнее показать, что (и как) то, чем занимаются математики, сидящие в различных комнатах, в действительности взаимосвязано, но избежать ощущения такого единства довольно трудно.

Наша книга состоит из замкнутых глав. Каждая глава посвящена углубленному изложению одной темы. Дойдя до сотой страницы, читателю вовсе не требуется знать, что написано на первой. Начинать чтение книги можно с любой главы, название или рисунки к которой вам больше понравятся. Даже если какая-нибудь тема излагается кратко, это отнюдь не означает, что следует ограничиваться поверхностными рассуждениями и трюизмами. Читателю можно помочь взобраться на вершины и опуститься в недра математики, не требуя, чтобы он непременно становился альпинистом или шахтером. Разумеется, для таких восхождений и спусков подходит далеко не каждый предмет. Значительную часть математики нельзя понять, не вдаваясь в технические подробности. Математикам это хорошо известно, но им также известно, что более глубокое понимание достигается именно в тех случаях, когда техники удается избежать. Математики всегда встремятся исключить технику, если это способствует лучшему пониманию. Ныне неспециалисту можно рассказать о математике гораздо больше и на более высоком научном уровне, чем пятьдесят лет назад. Подборку именно таких математически содержательных и в то же время достаточно доступных тем мы и предлагаем читателю. От него требуется лишь сохранить в своей памяти наиболее существенное из того, чему его учили в школе на уроках математики, и забыть все второстепенное, чтобы оно ему не мешало.

Математика в ходе своего развития переросла не только собственные границы, но и границы между своими отдельными областями. Если раньше с некоторым трудом еще удавалось указать границу между чистой и прикладной математикой, между геометрией, алгеброй и анализом, то ныне никто не отважится сказать, где кончается один раздел математики и начинается другой. Независимо от того, кто ставит задачу математику, лингвист или биолог, после того, как найдена ее математическая формулировка, она оказывается в самом центре чистой математики. Геометрические задачи в наше время решают алгебраическими методами, а алгебраические — геометрическими. Шаг от переключательной схемы вычислительной машины до схемы строения Вселенной и схемы нейронной сети, спрятанной под сводом нашего черепа, не столь уж велик! Эта неожиданно близкая связь между, казалось бы, отдаленными предметами показана и в нашей книге, и да не убоится читатель, если в одной и той же главе речь пойдет о кофейных мельницах и о далеких мирах.

В первой главе мы займемся измерением мира. Сначала — маленького мира, в котором мы живем, нашей Земли, затем большего мира, вмещающего в своих пределах Солнце и Луну, и наконец всей огромной Вселенной. Сначала — мира, который считали плоским, затем искривленного мира, обитателям которого приходится пользоваться совсем иной геометрией, чем та, которую они изучали в школе.

Во второй главе математик умудряется представить себе гостиницу с бесконечным числом комнат, сравнить бесконечные множества, доказать за один присест бесконечно много математических утверждений и выразить нескончаемую цепочку численных выкладок в виде конечных формул.

В третьей главе мы перекидываем мост от того, как записывали числа жители Древнего Вавилона и Египта, к человеческому мозгу, придумавшему эти числа и многое другое, и между прочим заводим речь об играх, допускающих строгую законченную теорию, и об азартных играх, о вычислительных машинах, думающих машинах, машинах, играющих в шахматы и занимающихся переводом, о том, что могут и чего не могут делать машины.

Особый мир — мир биологической клетки — открывается перед читателем в четвертой главе. Как клетка переводит наследственную информацию с языка дезоксирибонуклеиновых кислот на язык синтезируемых ею протеинов? Задача изящная, но бесполезная, ибо природе нет дела до того, что может выдумать математик.

В иные миры вводит читателя пятая глава. Мера перестает играть там привычную нам роль. Это миры, созданные воображением плохих художников, не умеющих провести прямую и начертить окружность, миры, карты которых следует мысленно представлять себе в виде кубов с причудливо склеенными точками наружной поверхности. В этой главе мы расскажем, что останется, если подвергнуть фигуры и отображения непрерывной деформации, а в заключение построим замечательную кривую, покрывающую весь квадрат.

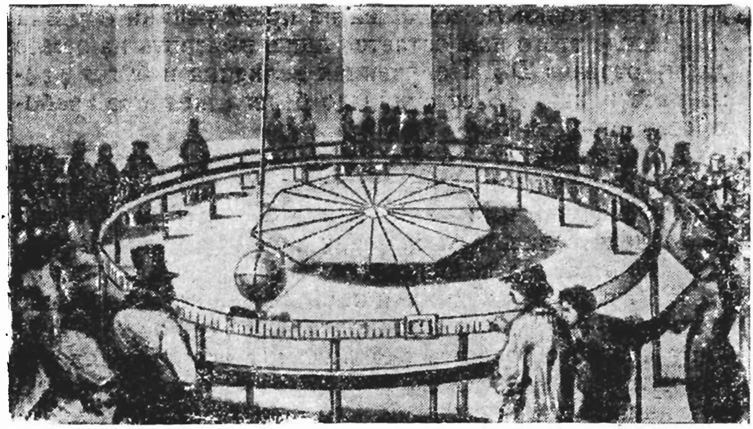
Шестая глава открывается гордым высказыванием Архимеда, который вознамерился перевернуть мир. В ней рассказывается о правиле рычага, о том, как оно проявляется в повседневной жизни и в коромысле весов. Затем от конкретных гирь мы перейдем к абстрактным грузам и будем нагружать ими коромысло воображаемых весов и невесомую плоскость, после чего снова вернемся к столь конкретным вещам, как определение оптимальной рецептуры шоколада и наиболее разумного расписания стирки скатертей и салфеток в ресторане.

В последней главе речь пойдет об озере, в котором отражается окрестный пейзаж, и о лунном лике, переходящем в себя при отражении, затем об отражениях на плоскости и в пространстве, о том, что можно сделать из отражений, о винтах, штопорах, радиоактивных атомах кобальта, позволяющих задать ориентацию пространства. Заканчивается седьмая глава далекими мирами, в которых штопоры и атомы кобальта могут быть зеркальными отражениями земных штопоров и атомов кобальта.

Быть может, у читателя возникнет желание упрекнуть меня за то, что время от времени я заставлял его преодолевать слишком большие трудности. В действительности же он должен быть признателен мне, что я никогда не делал его путь слишком легким.

Глава 1

Человек измеряет Вселенную



Геодезия

Первым ученым, чье имя дошло до нас, был Фалес Милетский (VI в. до н. э.). До него на протяжении двух тысячелетий наука оставалась безымянной. По преданию, Фалес предсказал солнечное затмение (этому он, по-видимому, научился у вавилонских астрономов). До нас дошли исследования Фалеса, относящиеся главным образом к геометрии. Так, например, утверждают, что Фалес Милетский умел измерять расстояние от судна, находящегося в море, до берега. Вот как выглядит метод Фалеса в изложении одного древнеримского геодезиста.

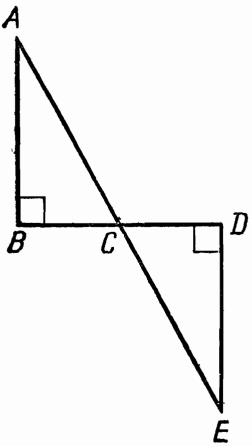


Рис. 1.

Предположим, что судно находится в точке A (рис. 1), а берег тянется вдоль отрезка BC, причем отрезок AB пересекает в точке B отрезок BC под прямым углом. В точке C вбит колышек. Продолжим отрезок BC за точку C так, чтобы $BC = CD$. Из точки D будем двигаться в сторону суши под прямым углом к отрезку CD до тех пор, пока не достигнем точки E, лежащей на продолжении отрезка AC. После этого нам остается лишь измерить на земле длину отрезка DE. Полученная величина и будет равна расстоянию AB от судна до берега, которое требовалось определить.

Не исключено, что Фалес побывал в Египте и поразил египтян, измерив высоту пирамид по длине отбрасываемой ими тени: для этого ему достаточно было действовать так же, как при определении расстояния от судна до берега.

В VI в. до н. э. Эвпалин построил на острове Самос водопровод для царя Поликрата. Среди прочего Эвпалину пришлось решить сложную инженерную задачу: прорыть тоннель сквозь гору Кастро. (Этот тоннель сохранился до наших дней. Его длина достигает километра, а ширина и высота — двух метров.) Проходку тоннеля вели одновременно с двух сторон. Ошибка в середине (в том месте, где должны были сойтись обе половины тоннеля) составила 10 м по горизонтали и 3 метра по вертикали. Это небольшая погрешность — относительная ошибка составила всего лишь 1%. Строители Силоамского тоннеля близ Иерусалима, созданного за 700 лет до н. э. при царе Гискии, действовали гораздо более грубыми методами и достигли существенно меньшей точности.

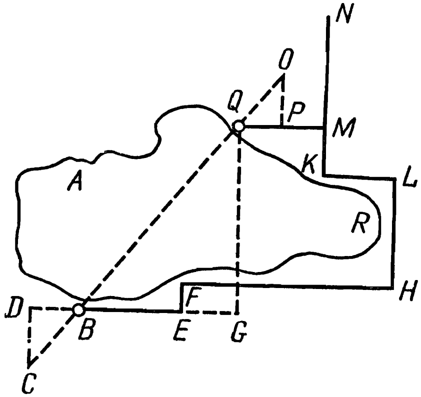


Рис. 2.

О том, каким образом Эвпалин мог построить свой тоннель, можно прочитать у Герона (правда, жившего гораздо позднее).

Предположим, что в горе $ABRQ$ (рис. 2) требуется пробить тоннель, ведущий от точки B к точке Q. Прежде всего обойдем вокруг горы по ломаной $BEFHLKMQ$, смежные отрезки которой расположены под прямым углом друг к другу. Измерив длину всех звеньев ломаной, нетрудно узнать длину непосредственно не измерявшихся отрезков BG и GQ. Перейдем затем в точку B и построим прямоугольный треугольник BDC с прямым углом при вершине D так, чтобы катет BD лежал на продолжении отрезка BE за точку B, а отношение катетов $DC:DB$ было равно отношению уже известных отрезков $GQ:BG$. Гипотенуза BC указывает направление, в котором следует вести проходку тоннеля из точки B. Аналогичное построение проведем и в точке Q.

Современный геодезист (если отвлечься от малосущественных деталей) поступил бы точно так же. Правда, он не стал бы строить треугольники DBC и OPQ. Поскольку отношение $GQ:BG$ ему известно, то тем самым он знает тангенс угла QBG и, следовательно, может найти по таблице тангенсов сам угол QBG. Современному геодезисту нет необходимости пользоваться только геометрическими построениями: неизвестный угол QBG он может восстановить и численными методами. Кроме того, при построении ломаной, ведущей в обход горы, современный геодезист не обязательно должен рассматривать лишь такие ломаные, у которых смежные звенья образуют друг с другом прямые углы: располагая таблицами логарифмов, он с не меньшей легкостью справится и с любыми другими углами.

Триангуляция

Искусство геодезистов состоит в том, чтобы определять расстояния и углы, не доступные непосредственному измерению. Измерять расстояния мерным жезлом или мерной лентой — занятие не из легких, гораздо проще измерять углы теодолитом. Представим себе, что геодезисту нужно определить расстояние между точками A и B (рис. 3), которое нельзя измерить непосредственно. Геодезист измерит длину отрезка AC и углы ACB и CAB. Зная сторону AC и два прилежащих к ней угла, он сможет определить все элементы треугольника ABC, в том числе и длину стороны AB. Если говорить более подробно, то длину стороны AB геодезист вычислит, воспользовавшись теоремой синусов

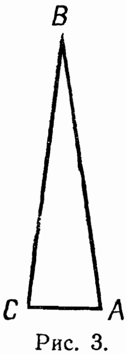


Рис. 3.

$$AB : AC = \sin(\angle ACB) : \sin(\angle ABC).$$

Длина отрезка AC и угол ABC известны непосредственно из измерений, $\angle ABC = 180° - \angle BCA - \angle BAC$. Имея под рукой таблицы синусов, современный геодезист без труда вычислит длину отрезка AB.

Если требуется измерить большое пространство, то его покрывают треугольниками (рис. 4), в которых измеряют все углы. Длину при этом приходится измерять лишь один раз, зато число непосредственно измеряемых углов превосходит число углов, величина которых первоначально неизвестна.

Например, может случиться, что в каком-нибудь треугольнике геодезист измеряет независимо друг от друга все три угла. Их сумма, как известно, должна

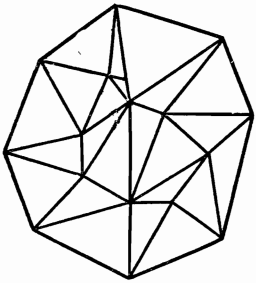


Рис. 4.

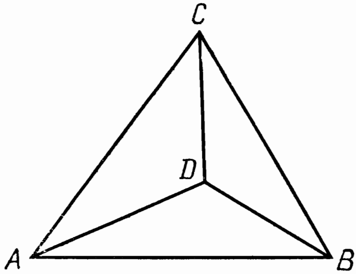


Рис. 5.

быть равна 180°, но в действительности, сложив три полученные путем измерения величины углов, мы, как правило, не получим в точности 180°. Специальный раздел общей теории ошибок — теория уравнивания — учит, каким образом следует распределять ошибку среди углов треугольника.

Классической задачей геодезии служит так называемая обратная засечка: внутри известного треугольника требуется построить точку D по измеренным углам, под которыми из нее видны стороны треугольника (рис. 5). Тригонометрические формулы, выражающие решение этой задачи, имеют довольно сложный вид.

Размеры Земли

До сих пор мы действовали так, будто поверхность Земли плоская. Именно таким образом представляли себе форму Земли люди в глубокой древности: нечто вроде диска, плавающего в океане. Подобных взглядов придерживался еще Фалес Милетский, о котором мы уже упоминали. Насколько можно судить, шарообразность Земли была обнаружена в VI или в V в. до н. э. Предание приписывает честь этого открытия Пифагору, имя которого незаслуженно носит известная теорема о соотношении сторон прямоугольного треугольника. Некоторые считают, что шарообразность Земли открыл Парменид, но имеются основания

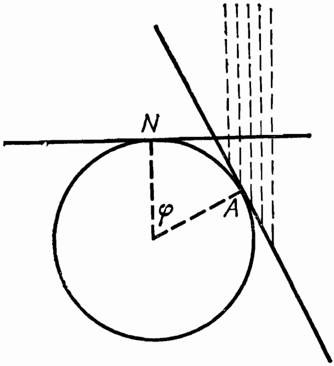


Рис. 6.

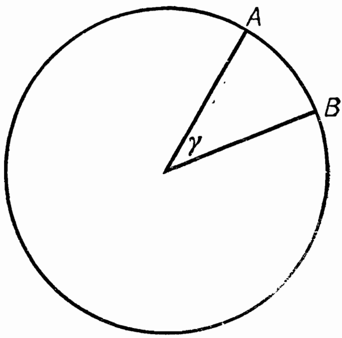


Рис. 7.

полагать, что еще юный Демокрит оставался в неведении относительно истинной формы земной поверхности.

К заключению о шарообразности Земли люди пришли, заметив, что при лунных затмениях Земля отбрасывает тень, имеющую форму круга. На мысль о сферичности земной поверхности наводил и повседневный опыт мореплавателей: известно, что когда вдали показывается корабль, то сначала из-за горизонта появляются верхние концы мачт и лишь затем расположенные ниже реи и прочие детали судна. Объяснить такие наблюдения удается, лишь приняв предположение о том, что Земля имеет форму шара.

Точнее всего форму Земли можно установить, воспользовавшись *астрономическими наблюдениями.* Наблюдатель, находящийся на поверхности Земли в точке A, видит небо иначе, чем наблюдатель, находящийся на Северном полюсе N (рис. 6). Плоскость которую наблюдатель, находящийся в точке A, считает горизонтальной, представляет собой касательную плоскость, проведенную к земной поверхности в точке A. Эта плоскость пересекает касательную плоскость, проведенную к земному шару на Северном полюсе, под углом φ. Например, если географическая широта точки A составляет $52°$, то $\varphi = 90° - 52° = 48°$. На Северном полюсе наблюдатель увидит Полярную звезду прямо над головой — в зените. Перейдя в точку A, наблюдатель обнаружит, что Полярная звезда склонилась к горизонту, отойдя от вертикали на угол φ по направлению к северу. К северу сместятся и остальные небесные тела, причем все на один и тот же угол φ. Таким образом, взглянув на небо, можно узнать, какова географическая широта того места, где мы находимся. Если переместиться на γ градусов к югу, двигаясь строго по меридиану, например, перейти из точки A в точку B (рис. 7), то в северной части неба светила опустятся на угол γ, а в южной — на столько же градусов поднимутся. Если определить угол γ (он «написан» на небе, нужно лишь уметь прочитать его) и измерить на Земле дугу AB, то нетрудно вычислить длину окружности Земли. Действительно,

$$\frac{\textit{Длина окружности Земли}}{\textit{Длина дуги } AB} = \frac{360°}{\gamma}.$$

Эта идея очень проста, и она вполне могла прийти в голову тому, кто открыл шарообразность Земли. Измерить угол на небе не составляло труда, гораздо труднее было измерять расстояния на Земле. Правда, расстояния между портовыми городами можно было измерять по тому, сколько дней должен находиться в пути корабль, идущий из одного порта в другой, но кто знал скорость корабля? И как было установить, лежит ли один город строго южнее другого?

Измерение окружности Земли, оставшееся непревзойденным на протяжении двух тысячелетий, произвел в III в. н. э. хранитель знаменитой Александрийской библиотеки математик, поэт, филолог и географ Эратосфен. Из рассказов путешественников Эратосфену было известно, что город Сиена (ныне Асуан), расположенный у первых нильских порогов, находится строго к югу от Александрии и что в Сиене имеется отвесный колодец, дно которого ровно в полдень самого длинного дня в году освещено Солнцем. Следовательно, Сиена должна лежать на широте тропика Рака. В тот же день и в тот же час Солнце в Александрии отстоит от зенита на $^1/_{50}$ окружности. Таким образом, длина окружности Земли равна расстоянию от Александрии до Сиены, умноженному на 50. Это расстояние измерил для Эратосфена придворный «счетчик шагов». Оно оказалось равным 5 тыс. стадий, а окружность Земли — 250 тыс. стадий. Чему равнялась использованная Эратосфеном единица длины — стадия, нам не известно. Впрочем, и без этого ясно, что полученное им значение длины окружности Земли не могло быть точным: предположение о том, что Сиена расположена строго к югу от Александрии, не соответствовало действительности. Кроме того, подсчет числа шагов отнюдь не принадлежит к точным методам измерения расстояний, и округление числа стадий до 5 тыс. также внесло свою ошибку. Если считать, что стадия равна 160 м, то получается правильное значение длины земной окружности — 40 тыс. км, но коль скоро пределы ошибки неизвестны, то само по себе такое число еще ни о чем не говорит.

В начале XVII в. Землю измерили снова. На этот раз попытку определить размеры нашей планеты предпринял голландец Виллеброрд Снеллиус. В окрестностях Лейдена он тщательно измерил отрезок (AB) прямой — базис длиной около 1 км. Из его концов Снеллиус мысленно провел прямые к шпилю лейденской ратуши C и колокольне кирхи D в одной из расположенных поблизости деревень и измерил углы CAB, DAB и CBA, DBA, что позволило ему вычислить расстояние CD. К пунктам C и D Снеллиус, произведя соответствующие измерения, присоединил координаты самых высоких зданий в ближайших городах и деревнях и постепенно добрался до Алкмара на севере и Бергена-оп-Зома на юге. Но и это было еще не все. И в Алкмаре, и в Бергене-оп-Зоме Снеллиус определил высоту Северного полюса мира и лишь тогда, пользуясь результатами своих измерений, приступил к вычислению длины окружности земного шара. Если полученную им величину выразить в современных единицах длины, то она окажется равной 38 650 км. Иначе говоря, относительная ошибка составляет 4%. Это очень большая погрешность, если учесть, что произведенные Снеллиусом геодезические измерения отличались высокой точностью. По-видимому, при определении высоты Северного полюса Снеллиус пользовался неисправным инструментом.

Следуя принципам, заложенным Снеллиусом, французы на протяжении двух столетий производили все более точные измерения Земли. Их инструменты были снабжены оптическими трубами. Во время этих измерений французским геодезистам удалось установить и сплюснутость Земли. Когда во времена Французской революции решили положить конец неразберихе, царившей в области мер и весов, то за единицу длины условились принять метр — одну десятимиллионную расстояния, измеренного вдоль меридиана от Северного полюса до экватора. Длина окружности Земли при этом выражается круглым числом — 40 000 км. Чтобы точно определить новую единицу длины и установить ее эталон, пришлось вновь заняться измерением длины окружности земного шара. Градусные измерения производились от Дюнкерка до Барселоны и длились 6 лет. Столь грандиозное предприятие не обошлось без конфуза: один из геодезистов искусственно «приукрашал» результаты своих измерений и тем самым вносил в них ошибку. Когда позднее все данные были тщательно проверены, принятая единица длины — метр — по различным причинам оказалась слишком короткой: расстояние между полюсом и экватором, выраженное в новых единицах, оказалось равным не 10 000 км, а на 2,3 км больше. Эталонный стержень, более столетия служивший образцовым метром, с соблюдением всех предосторожностей хранится в особом погребе неподалеку от Парижа. Ныне метр обрел независимость от эталонного стержня: теперь, чтобы получить длину 1 м, в качестве эталона используют длину волны определенным образом выбранного света. С недавних пор появилась возможность находить фигуру Земли гораздо более точно, чем это некогда можно было сделать геодезическими методами: для проведения высокоточных измерений стали использовать искусственные спутники. Расстояние от спутника до поверхности Земли определяют с помощью радара. Стоит лишь заметить, сколько времени проходит с момента посылки электромагнитной волны до возвращения ее эха, чтобы сразу стало известно, на каком расстоянии находится спутник от станции наблюдения. Зная расстояния в любой момент времени, нетрудно вычислить траекторию спутника, которая зависит от действия силы земного тяготения и формы Земли. По виду этой траектории можно судить о форме нашей планеты.

Астрономические расстояния

«Человек измеряет Вселенную» — таково название этой главы. Вселенной, или миром, иногда называют и Землю, но теперь мы будем вкладывать в это понятие гораздо более широкий смысл. И эту, более широкую Вселенную люди также пытались измерить еще в глубокой древности.

В III в. н. э. Аристарх Самосский, предшественник Коперника по созданию гелиоцентрической картины мироздания, написал труд под названием «Расстояния и размеры Солнца и Луны». В своем труде Аристарх Самосский исходил из шести предположений, которые в современном виде выглядят следующим образом:

1. Луна светит отраженным светом, источником которого служит Солнце.
2. Земля находится в центре траектории, описываемой Луной.
3. Когда Луна кажется нам разделенной тенью пополам, то окружность на лунной поверхности, служащая границей между светом и тенью, лежит в плоскости, проходящей через глаз наблюдателя.
4. Когда Луна кажется нам разделенной тенью пополам, то угол между направлениями Земля — Луна и Земля — Солнце составляет 87°.
5. Тень, отбрасываемая Землей, может покрыть небесное тело, радиус которого вдвое больше радиуса Луны.
6. Видимая ширина лунного диска равна 2°.

Затем следуют утверждения Аристарха.

1. Отношение расстояния Земля — Солнце к расстоянию Земля — Луна заключено между 18 : 1 и 20 : 1.

2. Отношение диаметра Солнца к диаметру Луны также заключено между 18 : 1 и 20 : 1.

3. Отношение диаметра Солнца к диаметру Земли заключено между 19 : 3 и 43 : 6.

Свои утверждения Аристарх доказывает со всей тщательностью, пользуясь евклидовой геометрией.

Первое утверждение по существу следует из четвертого предположения. Пусть Земля находится в точке A, а Луна — в точке B и Солнце — в точке C (рис. 8). Когда нам кажется, что Луна разделена тенью пополам, то угол при вершине B в треугольнике ABC прямой. Поскольку угол при вершине A в этот момент равен 87°, то угол при вершине C равен 3° и, следовательно,

$$\sin 3^\circ = \frac{AB}{AC}.$$

Пользуясь таблицами синусов, мы без труда можем убедиться в том, что отношение расстояния Земля — Солнце к расстоянию Земля — Луна заключено в пределах, указанных в первом утверждении Аристарха (необходимо лишь ввести поправку на размеры Земли).

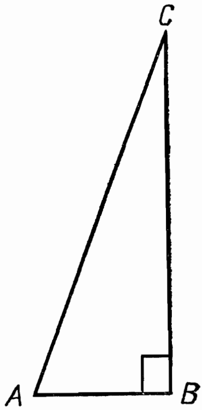


Рис. 8.

Далее Аристарх использует в своих рассуждениях известный из астрономических наблюдений факт, о котором он не упоминает в своих предположениях: при полном солнечном затмении Луна в точности покрывает Солнце. Отсюда следует (рис. 9), что диаметр Солнца относится к диаметру Луны так же, как расстояние Земля — Солнце к расстоянию Земля — Луна.

Пятое предположение Аристарха Самосского относится к лунным затмениям (рис. 10). Пусть A, B, C — центры Земли, Луны и Солнца, r_A, r_B, r_C — радиусы Земли, Луны и Солнца, а P — вершина конуса тени, отбрасываемой Землей. Как видно из рис. 10,

$$CP : AP : BP = r_C : r_A : 2r_B. \qquad (1)$$

Отношение

$$r_C : r_B = AC : AB$$

известно: мы вычислили его ранее. Обозначив его

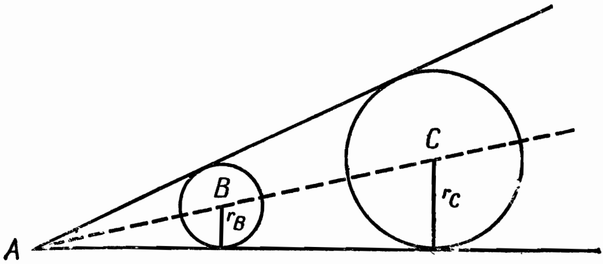


Рис. 9.

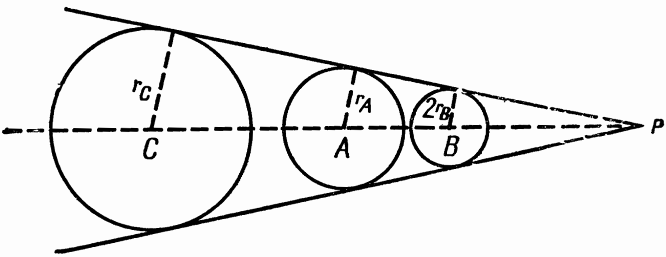


Рис. 10.

через p, получим

$$AC : AB = p. \qquad (2)$$

Соотношение (1) в новых обозначениях примет вид

$$CP : BP = \frac{1}{2} p. \qquad (3)$$

Соотношение (2) позволяет выразить длину отрезка AC через длину отрезка AB и известную величину p:

$$AC = p \cdot AB.$$

Таким образом,

$$BC = AC + AB = (p + 1)\, AB. \qquad (4)$$

Аналогично из соотношения (3) получаем

$$CP = \frac{1}{2} p \cdot BP, \tag{5}$$

откуда

$$BC = CP - BP = \left(\frac{1}{2} p - 1\right) BP. \tag{6}$$

Сравнивая соотношения (4) и (6) и вводя новое обозначение

$$\frac{p+1}{\frac{1}{2} p - 1} = q, \tag{7}$$

приходим к новому соотношению

$$BP = q \cdot AB, \tag{8}$$

в силу которого

$$AP = BP + AP = (q + 1) AB. \tag{9}$$

Наконец, из соотношений (5) и (8) мы находим

$$CP = \frac{1}{2} pq \cdot AB, \tag{10}$$

а из соотношений (9) и (10) —

$$CP : AP = \frac{1}{2} pq : (q + 1).$$

Отношение $CP : AP$ равно отношению радиуса Солнца к радиусу Земли, о котором говорится в третьем утверждении Аристарха. Подставляя найденное ранее значение p и соответствующее ему значение q, вычисляемое по формуле (7), в правую часть последнего равенства, мы убеждаемся, что отношение $r_C : r_A$ лежит в пределах, указанных Аристархом.

Поскольку отношения $r_C : r_A$ и $r_C : r_B$ теперь известны, ничто не мешает нам вычислить отношение $r_A : r_B$. Более того, используя шестое предположение Аристарха Самосского, можно найти отношение $AB : r_B$ (рис. 11) и, следовательно, $AB : r_A$, то есть расстояние Земля — Луна, измеренное в радиусах Земли.

Если говорить о численных значениях радиусов Солнца и Луны и расстояний Земля — Луна и Земля — Солнце, полученных Аристархом, то они оказались весьма неточными. Причина столь больших расхождений с истинными значениями кроется главным образом в четвертом предположении. Угол, о котором говорится в этом предположении, в действительности отличается от 90° гораздо меньше, чем полагал Аристарх (по-видимому, сам Аристарх не измерял его). Расстояние Земля — Луна, измеренное в радиусах Земли, было определено позднее (также в глубокой древности) с весьма высокой точностью. Наоборот, если речь заходила о расстоянии Земля — Солнце и, следовательно, о величине радиуса Солнца, то вплоть до нового времени приводили сильно заниженные значения. Ныне расстояния до «близких» объектов во

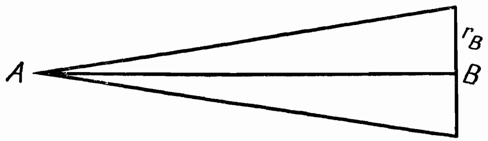


Рис. 11.

Вселенной можно весьма точно измерить методами радиолокации.

С появлением и усовершенствованием телескопа и подзорных труб измерения расстояний и на Земле, и во Вселенной стали производиться со все возрастающей точностью: ведь, по сути дела, расстояния до ближайших небесных тел можно определять теми же методами, что и расстояния на Земле. Пусть, например, A и B — две точки на Земле, а C — точка на Луне (рис. 12). Для того чтобы вычислить расстояния AC и BC, необходимо лишь измерить расстояние AB и углы при вершинах A и B треугольника ABC. По мере удаления точки C от поверхности Земли стороны AC и BC становятся почти параллельными, а когда точка C удалится на очень большое расстояние, неотличимыми от параллельных. Определить, как далеко от Земли отстоит точка их пересечения, становится невозможным. Поэтому мы не можем отличить очень большие расстояния друг от друга и от бесконечно большого расстояния. Где проходит граница «бесконечности», зависит от чувствительности инструмента, которым мы измеряем углы при вершинах A и B. Все зависит от величины $\angle ACB = 180° - \angle CAB -$

— $\angle CBA$. Если углы CAB и CBA мало отличаются от прямых, то соотношение

$$AC : AB = \sin(\angle ABC) : \sin(\angle ACB)$$

можно приближенно заменить соотношением

$$AC = \frac{AB}{\sin(\angle ACB)}.$$

Если мы хотим воспользоваться этим соотношением для того, чтобы определить расстояние AC, то угол

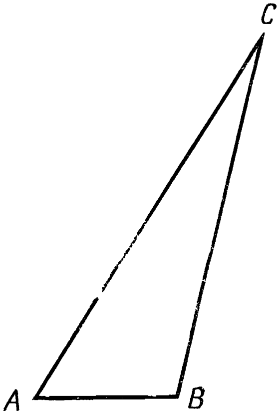


Рис. 12.

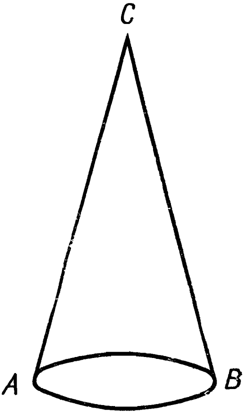


Рис. 13.

ACB не должен быть исчезающе малым: наш измерительный прибор должен отличать его от 0°.

Разумеется, «горю» можно помочь, выбрав побольше базис AB. Если мы ограничим свои поиски нашей планетой, то самый большой базис, какой только можно построить на Земле, не превосходит ее диаметра (около 12 700 км). По космическим масштабам это исчезающе малое расстояние. Но даже оставаясь на Земле, мы все же можем построить гораздо больший базис: Земля обращается вокруг Солнца, и расстояние между наиболее удаленными друг от друга точками ее траектории составляет около 300 млн. километров. Правда, из двух таких точек траектории Земли наблюдатель видит удаленную точку C не одновременно, поскольку Земле требуется полгода, чтобы переместиться из одной точки в другую. Угол при вершине C можно определить: $\angle ACB = 180° - \angle CAB - \angle CBA$. Половина угла ACB называется *параллаксом* точки C. Именно под таким углом видел бы радиус земной орбиты наблюдатель, находись он в точке C. Расстояние, на которое должен удалиться наблюдатель, чтобы видеть радиус орбиты Земли под углом в одну секунду, астрономы условились называть *парсеком*. Полная окружность содержит $360 \cdot 60 \cdot 60 = 1296 \cdot 10^3$ угловых секунд. Синус угла в одну секунду с достаточной точностью можно считать равным дуге, стягивающей центральный угол в одну секунду, то есть $2\pi/1296 \cdot 10^3$. Радиус траектории Земли составляет около $150 \cdot 10^6$ км. Следовательно, один парсек равен примерно

$$\frac{1296 \cdot 10^3 \cdot 150 \cdot 10^6}{2\pi} \approx 31 \cdot 10^{12} \text{ км}.$$

Итак, один парсек составляет 31 млрд. км, или, если за единицу длины принять световой год, $^3/_4$ светового года.

Ближайшие после Солнца неподвижные звезды находятся от нас на расстоянии уже в 4 световых года, поэтому их параллаксы очень малы. Это обстоятельство доставило множество хлопот Копернику и Галилею. Сторонники геоцентрической системы говорили: «Если Земля обращается вокруг Солнца, то неподвижные звезды должны обладать ненулевым параллаксом, однако всем известно, что это не так». Разумеется, сторонники гелиоцентрической системы могли возразить им: «Неподвижные звезды находятся от нас на столь большом удалении, что их параллакс не поддается измерению». Но подобный ответ противники гелиоцентрической системы считали отговоркой. Даже изобретение телескопа не позволило дать более убедительный ответ. Параллаксы неподвижных звезд удалось впервые измерить лишь в XIX в. Методы измерения параллаксов все время совершенствуются, но даже лучшие из них позволяют измерять расстояния, не превышающие сто световых лет.

Для измерения Вселенной на расстояниях, превышающих возможности звездной «геодезии», используют другие методы. Отношение видимых яркостей, или, как принято теперь называть, блеска, двух объектов одной и той же истинной яркости совпадает с обратным отношением квадратов расстояний до них. Следовательно, расстояния до двух объектов можно сравнивать, если известно, что абсолютная яркость этих объектов одинакова. Но каким образом это можно установить? Ведь существуют звезды различных типов и классов, отличающиеся не только блеском, но и цветом (точнее, спектром излучаемого ими света), и, что особенно важно, среди звезд нередко встречаются так называемые переменные звезды, блеск которых изменяется со временем. Наблюдая за «ближними» звездами, удалось установить, что звезды одного и того же типа обладают одинаковой абсолютной яркостью. Разумно предположить, что и более удаленные от нас звезды одного и того же типа также обладают одинаковой абсолютной яркостью. Приняв такую гипотезу, мы получаем возможность судить о расстояниях до звезд по их блеску. Аналогичным образом астрономы поступают и с отстоящими от нас на огромные расстояния млечными путями (спиральными туманностями): относительно их также предполагают, что они обладают приблизительно одинаковой абсолютной яркостью. Даже разбегание спиральных туманностей помогает нам измерять Вселенную. Наша Вселенная расширяется, и поэтому спиральные туманности разбегаются со скоростью, пропорциональной расстоянию, отделяющему туманность от нас. Скорость разбегания можно измерить по сдвигу спектральных линий и таким образом оценить расстояние, отделяющее нас от спиральных туманностей. Оно составляет несколько миллиардов световых лет.

Жизнь на искривленной поверхности

Но вернемся снова на Землю! Предположим, что ее обитателям неизвестна такая наука, как астрономия, а звезды всегда скрыты от взоров землян за толстым слоем облаков. Представим себе даже, что жители Земли не в силах вознести свои головы в просторы Вселенной столь далеко, чтобы, заметив постепенно скрывающиеся за горизонтом корабли, прийти к выводу об искривленности земной поверхности. Могли бы обитатели Земли все-таки открыть ее шарообразность? Разумеется, могли бы! Ведь тот, кто все время идет по поверхности Земли прямо, в действительности описывает окружность большого круга, то есть движется по границе круга, плоскость которого проходит через центр земного шара, и в конце концов возвращается в исходную точку. На плоской Земле совершить кругосветное путешествие, двигаясь все время прямо, было бы невозможно. Но представим себе, что кто-нибудь всю свою жизнь проведет в пути, двигаясь все время только вперед, и все же не успеет вернуться в свой родной город. Такой путешественник не вправе утверждать, будто Земля плоская: для завершения кругосветного путешествия человеческой жизни может и не хватить. Но удастся ли ему, не отдаляясь особенно от стен родного города, установить, плоской ли или искривленной является поверхность Земли, и определить, сколь сильна кривизна?

Предположим, что наш любознательный обитатель Земли — геометр. Он знает, что сумма углов в треугольнике равна двум прямым углам, и ему известна теорема Пифагора. Обе эти теоремы относятся к так называемой планиметрии, или геометрии плоскости. На сфере или на любой другой искривленной поверхности они не выполняются, хотя и на сфере и на любой другой искривленной поверхности существуют прямые линии. То, что обитатель сферической поверхности назвал бы прямой, — кратчайшую линию, соединяющую две точки на поверхности сферы, — мы называем дугой большого круга. Можно высказать и более общее утверждение: кратчайшие линии, которые обитатель той или иной искривленной поверхности назвал бы *прямыми*, нам, смотрящим на них извне, представляются в виде *кривых* (рис. 14). Из таких линий обитатель искривленной поверхности строит свои треугольники. Измерив стороны и углы этих треугольников, он может тем самым проверить, выполняются ли для поверхности, на которой он живет, теоремы планиметрии, а затем сделать соответствующие выводы.

Предположим, что наш геометр построил сферический треугольник ABC (рис. 15). Угол, который он измеряет между *своими* прямыми AB и AC (с нашей точки зрения, стороны AB и AC являются дугами окружностей), равен углу между плоскостями MAB и MAC (M — центр сферы). Измеренный угол заведомо больше *плоского* угла BAC, и сумма углов в сферическом треугольнике ABC больше суммы углов плоского треугольника, то есть больше двух прямых углов. Разность между суммой углов сферического

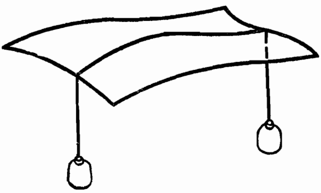


Рис. 14.

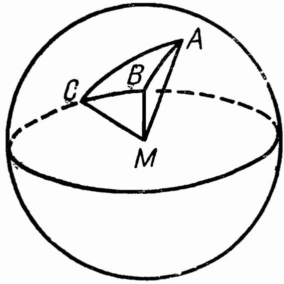


Рис. 15.

треугольника и 180° называется сферическим *избытком* данного треугольника. Чем больше сферический треугольник, тем больше его избыток. Например, выберем вершину A сферического треугольника на северном полюсе сферы, а вершины B и C — на экваторе так, чтобы их разделяла дуга в четверть окружности (рис. 16). Такой сферический треугольник содержит три прямых угла, поэтому сумма его углов равна 270°, а его сферический избыток — 90°.

Можно показать, что сферический избыток пропорционален площади сферического треугольника. То, что обитатель поверхности сферы называет прямой, для постороннего наблюдателя выглядит дугой большого круга, то есть круга, плоскость которого проходит через центр сферы. Окружности двух больших кругов пересекаются в точках, лежащих на противоположных концах диаметра сферы. Обе окружности разбивают поверхность сферы на четыре «двуугольника» (на рис. 17 один двуугольник заштрихован). Если угол двуугольника равен $\alpha \cdot 360°$, то его площадь составляет α-ю часть поверхности сферы. Обозначив площадь поверхности сферы через S, мы сможем представить площадь двуугольника в виде αS.

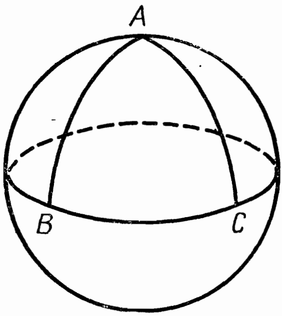


Рис. 16.

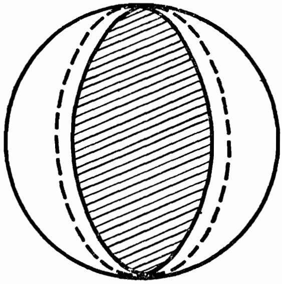


Рис. 17.

Рассмотрим на сфере треугольник ABC, ограниченный дугами больших кругов (рис. 18). Продолжим сторону AB за вершину B, сторону BC за вершину C и сторону CA за вершину A до тех пор, пока продолженные стороны не пересекутся еще раз. Пусть A' — точка пересечения дуг, проходящих через вершину A, B' — точка пересечения дуг, проходящих через вершину B, и C' — точка пересечения дуг, проходящих через вершину C (заметим, что точка A' служит антиподом точки A, то есть диаметрально противоположна точке A, точка B' — антипод точки B и точка C' — антипод точки C). Проведенные дуги разбивают поверхность сферы на пять частей:

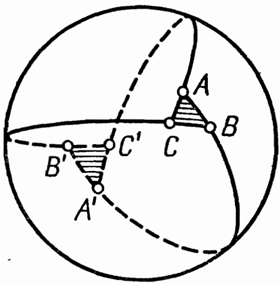


Рис. 18.

1) треугольник ABC;
2) треугольник $A'B'C'$;
3) двуугольник с вершинами в точках A, A' и сторонами, проходящими через точки B и C;
4) двуугольник с вершинами в точках B, B' и сторонами, проходящими через точки C и A';

5) двуугольник с вершинами в точках C, C' и сторонами, проходящими через точки A и B'.

Оба треугольника имеют одинаковую площадь, которую мы обозначим Δ. Если $\alpha\cdot 360^\circ$, $\beta\cdot 360^\circ$, $\gamma\cdot 360^\circ$ — внешние углы треугольника ABC, то площади двуугольников равны αS, βS, γS. Сумма площадей треугольников ABC, $A'B'C'$ и трех двуугольников должна быть равна площади полной поверхности сферы, то есть

$$2\Delta + (\alpha+\beta+\gamma) S = S.$$

Если $\alpha'\cdot 360^\circ$, $\beta'\cdot 360^\circ$, $\gamma'\cdot 360^\circ$ — внутренние углы треугольника, то $\alpha'\cdot 360^\circ + \alpha\cdot 360^\circ = 180^\circ$ и т. д., в силу чего $\alpha = {}^1/_2 - \alpha'$, $\beta = {}^1/_2 - \beta'$, $\gamma = {}^1/_2 - \gamma'$. Подставляя полученные выражения в выведенное выше соотношение, получаем

$$2\Delta + \left(\frac{3}{2} - \alpha' - \beta' - \gamma'\right)\cdot S = S,$$

откуда

$$\alpha' + \beta' + \gamma' - \frac{1}{2} = \frac{2\Delta}{S}.$$

Сферический избыток треугольника равен

$$(\alpha' + \beta' + \gamma')\cdot 360^\circ - 180^\circ = \left(\alpha' + \beta' + \gamma' - \frac{1}{2}\right)\cdot 360^\circ = \frac{2\Delta}{S}\cdot 360^\circ. \quad (1)$$

Таким образом, у сферического треугольника с тремя прямыми углами, покрывающего ${}^1/_8$ поверхности сферы, сферический избыток действительно оказывается равным 90°.

Представим теперь себя на месте обитателя сферической поверхности, не подозревающего об искривленности его мира. Мы измеряем малый треугольник и сначала ничего не замечаем. Затем переходим к измерению треугольника бóльших размеров, и тут выясняется, что сумма углов треугольника больше 180°. Первое объяснение, которое приходит в голову: в измерения вкралась ошибка. Но повторные измерения суммы углов треугольников вновь приводят к значениям, превышающим 180°, причем отклонения от 180° возрастают с увеличением треугольника. Как это объяснить? Систематическое исследование показывает, что избыток пропорционален площади треугольника. И тут вас осеняет, что вы живете на сфере. Более того, нетрудно вычислить и величину полной поверхности этой сферы: если S — площадь полной поверхности сферы, а Δ — площадь треугольника со сферическим избытком ε, то

$$S = \frac{2\Delta \cdot 360^\circ}{\varepsilon}.$$

Зная площадь поверхости сферы, без труда находим ее радиус r, поскольку

$$S = 4\pi r^2.$$

Величину

$$K = \frac{1}{r^2}$$

принято называть *кривизной* сферы. Чем меньше радиус сферы, тем сильнее она искривлена. Пользуясь результатами своих измерений, обитатель сферической поверхности может представить кривизну своего мира в виде соотношения

$$K = \frac{2\pi}{\Delta} \cdot \frac{\varepsilon}{360^\circ}.$$

Он без труда находит K по измеренным Δ и ε.

То же соотношение остается в силе и в том случае, когда мир представляет собой не идеальную сферу, а некоторую неправильную изогнутую поверхность. Для такого мира соотношение между избытком и площадью треугольника так же, как и кривизна, изменяется при переходе от точки к точке. Но и в таком «непостоянном» мире соотношение между избытком и площадью треугольника служит мерой местной (локальной) кривизны.

Определенная при помощи приведенного выше выражения кривизна может быть не только положительной или нулевой, но и отрицательной. Об отрицательной кривизне говорят в тех случаях, когда сумма углов треугольника меньше 180°. Отрицательный избыток называется *дефектом.* С отрицательной кривизной мы встречаемся не только в геометрии, но и в реальной жизни. Представим себе, что мы находимся в горах около перевала. Карта с линиями равной высоты в окружности перевала выглядит примерно так, как показано на рис. 19. Перевал находится на высоте 1000 м над уровнем моря, к востоку и к западу местность повышается, перевал идет в направлении с севера на юг. Треугольники, построенные в окрестности перевала, обладают дефектом (это утверждение мы приводим без доказательства) — кривизна местности вблизи перевала отрицательна.

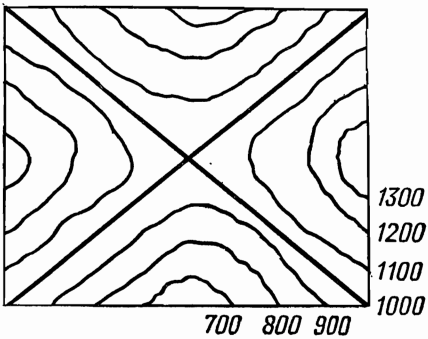


Рис. 19.

Разумеется, теорема о сумме углов треугольника — не единственная теорема планиметрии, которая перестает быть верной в искривленном мире. Вернувшись снова на сферу, мы обнаружим, что в сферической геометрии не выполняется теорема Пифагора. Сферический треугольник с тремя прямыми углами является равносторонним, но вместо обычной теоремы Пифагора на сфере мы приходим к соотношению $c^2 < a^2 + b^2$. По величине разности между правой и левой частью этого неравенства («дефекту» теоремы Пифагора) обитатель сферы также может вычислить кривизну своего мира.

Прежде чем рассмотреть еще одно отклонение от планиметрии, которое могут заметить обитатели сферического мира, займемся миром, имеющим форму конической поверхности.

Разрезав боковую поверхность конуса вдоль образующей, показанной на рис. 20 пунктиром, мы сможем развернуть его боковую поверхность на плоскость. Плоская развертка боковой поверхности конуса имеет вид кругового сектора с углом раствора φ. Следовательно, длина $2\pi\rho$ ограничивающей его дуги окружности относится к полной длине окружности $2\pi s$, как $\varphi : 360°$, то есть

$$\varphi = \frac{\rho}{s} \cdot 360°.$$

Развернув на плоскость («разгладив») боковую поверхность конуса, мы ничего не растянули и не сжали. Расстояния между любыми двумя точками остались

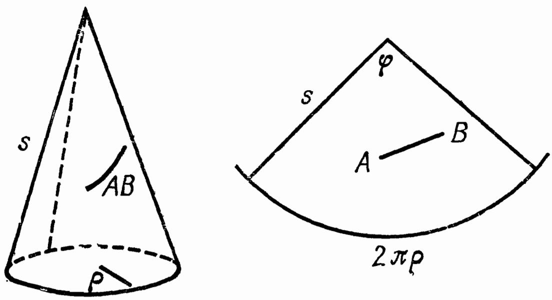


Рис. 20.

неизменными, и, следовательно, сохранилась вся геометрия поверхности. То, что на плоской развертке представляется в виде прямой, обитатели конической поверхности также называют прямой. Можно подумать, будто на боковой поверхности конуса действует та же геометрия, что и на плоскости, но такое предположение неверно. Если обитатель конического мира не совершает больших путешествий, то он и не замечает, что передвигается не по плоскости. Отправим теперь обитателя конической поверхности в кругосветное путешествие, попросив его захватить с собой стрелку, лежащую в плоскости развертки, и следить за тем, чтобы она все время оставалась параллельной самой себе. Пусть в исходной точке A маршрута стрелка направлена вдоль образующей конуса (рис. 21). Когда путешественник вернется в точку A (дойдет до точки A, изображенной на развертке слева), стрелка образует с проходящей через точку A

направляющей конуса угол φ. Нашему путешественнику не остается ничего другого, как констатировать, что стрелка, направление которой он поддерживал неизменным на протяжении всего кругосветного путешествия, повернулась на угол φ. При путешествиях на малые расстояния ничего подобного не происходит. Для того чтобы стрелка повернулась, необходимо обойти вокруг вершины конуса. Если путешественник обойдет вершину конуса не один, а несколько раз, то угол поворота стрелки φ увеличится пропорционально числу обходов вокруг вершины конуса.

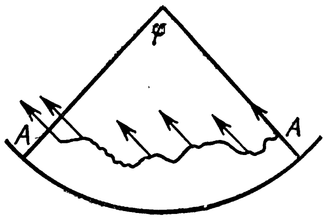


Рис. 21.

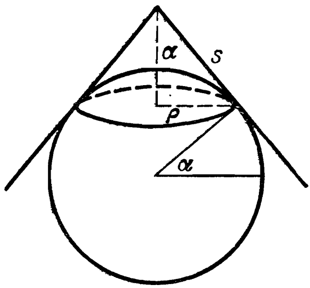


Рис. 22.

С точки зрения внешнего наблюдателя, стрелка поворачивается на угол $\varphi = (\rho : s) \cdot 360°$. С точки зрения самого путешественника, все выглядит несколько иначе. Внешний наблюдатель видит, что наш путешественник совершает вокруг вершины конуса полный оборот: описывает угол в 360°. Самому путешественнику кажется, что стрелка к концу пути поворачивается на угол

$$360° - \varphi = \left(1 - \frac{\rho}{s}\right) \cdot 360°.$$

Покинем теперь конический мир и возвратимся на поверхность сферы. Отправим ее обитателя в путешествие вокруг северного полюса по параллели и попросим его также захватить с собой стрелку и следить за тем, чтобы она все время оставалась параллельной самой себе (в смысле геометрии сферы). Каким образом может выполнить нашу просьбу «сферический» путешественник? Представим себе, что на сферу надет конус, касающийся ее вдоль той самой параллели, по которой наш путешественник обходит северный полюс (рис. 22). Когда обитатель сферической поверхности движется по параллели, ему приходится иметь дело лишь с непосредственной окрестностью параллели. В то же время параллельный перенос стрелки не зависит от того, как устроена сферическая поверхность по соседству с параллелью. У обитателя сферической поверхности стрелка будет вести себя так же, как у его собрата, живущего на конической поверхности, то есть на поверхности конуса, касательного к сфере. Вернувшись домой после кругосветного путешествия, обитатель сферы обнаружит, что стрелка отклонилась от первоначального направления.

Сколь велико это отклонение? Его величину мы вычислили, когда рассматривали кругосветные путешествия на конусе:

$$\left(1 - \frac{\rho}{s}\right) \cdot 360^\circ,$$

где ρ — радиус, а s — длина образующей конуса от вершины до исходной точки маршрута. Выразив эту величину через основные характеристики касательного конуса, мы увидим, что отношение ρ/s равно синусу угла, обозначенного α на рис. 22 и называемого также географической широтой той параллели, по которой обитатель сферы совершает обход северного полюса. Отклонение стрелки от первоначального направления зависит от географической широты параллели α и составляет

$$(1 - \sin \alpha) \cdot 360^\circ.$$

Эту формулу нетрудно проверить: на северном полюсе, географическая широта которого равна 90°, эффект исчезает. С точки зрения внешнего наблюдателя, путешественник вместе со стрелкой поворачивается на месте на 360°. Следовательно, после поворота стрелка образует со своим первоначальным положением угол в 360°. С точки зрения самого путешественника, со стрелкой ничего не происходит.

Когда путешественник обходит вокруг северного полюса по параллели, к северу от него остается сферическая шапочка — часть поверхности сферы, которую мы получим, разрезав сферу по параллели. Площадь сферической шапочки, лежащей выше параллели с географической широтой α, равна

$$(1 - \sin\alpha) \cdot 2\pi r^2,$$

где r — радиус сферы (эту формулу мы приводим без доказательства, считая ее известной). Измерив углы в радианах, а не в градусах, то есть положив полный

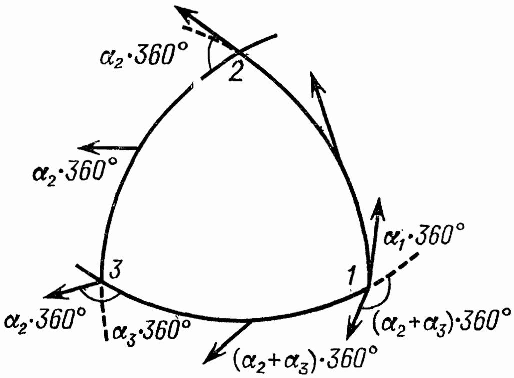


Рис. 23.

угол равным 2π вместо 360°, получим следующее соотношение:

$$\frac{\textit{Угол поворота стрелки}}{\textit{Площадь сферической шапочки}} = \textit{Кривизна}\left(= \frac{1}{r^2}\right).$$

Докажем это соотношение на другом примере. Обитатель сферы не считает параллель прямой. Его прямые в нашем понимании являются дугами окружностей больших кругов. Предположим, что живущий на поверхности сферы путешественник обходит треугольник, стороны которого кажутся ему прямыми: сначала из вершины *1* в вершину *2*, затем из вершины *2* в вершину *3* и, наконец, из вершины *3* снова в вершину *1* (рис. 23). Пока он движется по прямой, угол между направлением стрелки и направлением обхода остается неизменным. Предположим, что в исходной точке (вершине *1*) направление стрелки совпадает с направлением, в котором путешественник покидает вершину *1*. Тогда на протяжении всего перехода из вершины *1* в вершину *2* угол между направлением стрелки и направлением движения останется нулевым. В вершине *2* путешественник изменяет направление своего движения на угол $\alpha_2 \cdot 360°$, равный внешнему углу треугольника при вершине *2*. Разумеется, стрелка отнюдь не обязана следовать за изменением направления обхода и в момент, когда путешественник выходит из вершины *2* к вершине *3*, отстает от направления обхода, образуя с ним угол $-\alpha_2 \cdot 360°$. Таким образом, угол между направлением стрелки и направлением обхода остается до тех пор, пока путешественник не прибудет в вершину *3*. Здесь направление движения вновь изменяется (на этот раз на угол $\alpha_3 \cdot 360°$, равный внешнему углу треугольника при вершине *3*), и поэтому, когда путешественник отправляется по маршруту «вершина *3* — вершина *1*», направление стрелки образует с направлением обхода угол $(-\alpha_2 - \alpha_3) \cdot 360°$. Таким этот угол сохраняется вплоть до возвращения путешественника в вершину *1*. Дабы стать лицом в ту сторону, в которую он некогда вышел, направляясь к вершине *2*, путешественник должен повернуться на угол $\alpha_1 \cdot 360°$, равный внешнему углу треугольника при вершине *1*. При этом стрелка отстанет еще на угол $-\alpha_1 \cdot 360°$, и угол между ее направлением и направлением обхода стороны между точками *1* и *2*, то есть с первоначальным направлением стрелки, составит $(-\alpha_1 - \alpha_2 - \alpha_3) \times$ $\times\, 360°$. По причине, о которой мы уже говорили выше, прибавим к этому выражению 360° (или в радианах 2π). Тогда угол поворота стрелки при обходе треугольника, сторонами которого служат дуги больших кругов, можно представить в виде

$$(1 - \alpha_1 - \alpha_2 - \alpha_3) \cdot 360°.$$

По формуле (1) эта величина равна

$$\frac{2\Delta}{S} \cdot 2\pi,$$

где Δ — площадь треугольника, по периметру которого совершается обход, а S — площадь полной поверхности сферы, то есть $S = 4\pi r^2$. Разделив угол поворота стрелки на площадь треугольника, мы вновь получим $1/r^2$, то есть кривизну сферы.

Итак, обитатель сферической поверхности, как и обитатель всякого искривленного мира, может установить кривизну своего пространства, определив угол, на который повернется стрелка при обходе периметра какой-нибудь фигуры, и разделив этот угол на площадь фигуры. Вычисленная кривизна может быть не одинаковой для всех точек пространства и будет изменяться при переходе из одного места в другое.

Для параллельного переноса стрелки можно воспользоваться, например, волчком с горизонтальной осью (гироскопом). Из механики известно, что такой прибор сохраняет направление своей оси. Можно применить и маятник, поскольку направление его колебаний не изменяется. Разумеется, при переносе гирокомпаса или маятника в пространстве вдоль искривленной поверхности что-то все же должно изменяться: ось гироскопа и направление колебаний маятника лежат в касательной плоскости, а при путешествии по искривленной поверхности положение касательной плоскости в пространстве все время изменяется. Разумеется, те изменения, которые претерпевает направление оси гирокомпаса или направление колебаний маятника, происходят перпендикулярно мгновенному положению касательной плоскости. Касательные, или, как еще говорят, тангенциальные, составляющие обоих направлений остаются неизменными. Такой способ параллельного переноса стрелки вдоль искривленной поверхности (именно его мы и рассматриваем) носит название параллельного переноса в смысле Леви-Чивита в честь предложившего его в 1917 г. итальянского математика Т. Леви-Чивита. (Строго говоря, такой тип параллельного переноса был известен и раньше.)

В начале 1851 г. Париж облетела сенсация: все, кто еще сомневался, могли наглядно убедиться в том, что Земля вращается вокруг своей оси. Физик Фуко подвесил в Пантеоне маятник длиной 67 м с грузом 28 кГ. Совершая колебания, груз своим острием прочеркивал в песке, тонким слоем покрывавшем пол, следы, и каждый мог заметить, что направление колебаний за несколько минут успевает измениться. Парижане собственными глазами видели, как Земля поворачивается под маятником! На Северном полюсе острие маятника Фуко за 24 часа успело бы один раз начертить всю розу ветров, причем в направлении, противоположном вращению Земли. На более низких широтах маятник Фуко не столь проворен — к этому мы сейчас еще вернемся.

Опыт Фуко был первым поистине наглядным доказательством вращения Земли. Все другие эффекты, при помощи которых пытались доказывать вращение Земли предшественники Фуко, были столь слабыми, что заметить их можно было, лишь производя особо точные измерения. Для публичных демонстраций подобные эффекты не подходили. Опыт Фуко был сразу же повторен во многих странах мира. Маятник Фуко можно увидеть и в парижском Музее искусств и ремесел, и даже во Дворце Объединенных Наций в Нью-Йорке (подарок ООН от правительства Голландии) [1].

Но какое отношение имеет маятник Фуко к геометрии искривленных поверхностей? Самое непосредственное. Вращающийся вокруг собственной оси земной шар увлекает нас вместе с маятником Фуко. Если бы Земля была неподвижна, а мы двигались по ней с той же самой скоростью, с какой она увлекает нас в своем вращении, то за 24 часа мы описали бы полную окружность — обошли вокруг Северного полюса по параллели. Ранее мы уже установили, что при таком обходе стрелка, удерживаемая так, чтобы она все время оставалась параллельной самой себе, повернется относительно своего первоначального положения, и вычислили угол между начальным и конечным направлениями стрелки: $(1 - \sin\alpha)\cdot 360°$, где α — географическая широта той параллели, по которой совершается обход. Поскольку нас интересует угол отклонения между начальным и конечным положениями стрелки не на неподвижной, а на вращающейся Земле, то приведенный выше угол поворота

[1] Один из самых больших маятников Фуко (длиной 98 м) демонстрируется в Исаакиевском соборе в Ленинграде. — *Прим. перев.*

необходимо вычесть из 360°. Таким образом, плоскость колебаний маятника Фуко спустя 24 часа составит со своим начальным положением угол

$$\sin\alpha \cdot 360°.$$

Это означает, что на нашей широте плоскость колебаний маятника Фуко поворачивается примерно втрое медленнее часовой стрелки (успевающей за 24 часа *дважды* обойти циферблат). Если часовая стрелка имеет достаточно большую длину, то ее движение можно заметить невооруженным глазом. В опыте Фуко длине часовой стрелки соответствует довольно значительная величина отклонения огромного маятника от положения равновесия.

Кривизна нашего мира

Итак, обитатель некой искривленной поверхности, пользуясь лишь «внутренними» измерениями, вполне может установить, что живет не в плоском, а в искривленном мире, и даже вычислить его кривизну, для чего вовсе не нужно выходить за пределы этого мира. Говоря о мире, мы имеем в виду некую поверхность и считаем, что ее обитатель как бы прикован к ней незримыми цепями. В известной мере именно так обстоит дело с людьми, живущими на поверхности земного шара. И все же человек поднялся до таких высот (не столько в прямом, сколько в переносном смысле), что обнаружил шарообразность Земли еще до того, как смог судить о ней на основании своих измерений. Кроме того, собственное зрение убеждает человека в том, что Вселенную составляет не только та поверхность, по которой он может передвигаться.

Мы уже видели, как человек измеряет космическое пространство — не линейкой и не мерной лентой, ибо их оказывается недостаточно для измерения больших расстояний на Земле. Геодезист, желая убедиться, что три точки расположены на одной прямой, визирует их. Прямые, которые он при этом видит, представляют собой световые лучи — так же как и прямые, которые строят астрономы во Вселенной.

Еще геодезистам было известно, что предположение о прямолинейности световых лучей далеко не всегда соответствует действительности. Опущенное в воду весло лишь кажется сломанным. Причина мнимой «поломки» весла состоит в том, что при переходе из воздуха в воду световые лучи изменяют свое направление — *преломляются*. Геодезист, прежде чем принять допущение о прямолинейности световых лучей, вводит поправку на преломление света. То же делает и астроном, производя свои измерения Вселенной. Солнечные лучи, преломляясь в атмосфере Земли, искривляются, поэтому мы можем в течение непродолжительного времени наблюдать Солнце даже после того, как оно в действительности скрылось за горизонтом. При наблюдении светил вблизи горизонта астроному приходится вводить гораздо более существенную поправку на преломлене света, чем вблизи зенита.

Преломление — не единственная причина искривления световых лучей. Существует множество других причин, одна из них — сила тяготения. Световые лучи, идущие к нам от далекой звезды и проходящие вблизи Солнца, искривляются подобно траекториям комет. Земному наблюдателю кажется, что такая звезда несколько больше смещена в сторону от Солнца, чем в действительности. Это явление было предсказано Эйнштейном и полностью подтвердилось при наблюдении солнечных затмений.

Введя соответствующие поправки на искривление световых лучей, мы можем воспользоваться ими для построения треугольников в космическом пространстве. У «малых» треугольников (размеры которых сравнимы с размерами Солнечной системы) не удается обнаружить отклонения суммы углов от 180°: на таких расстояниях космическое пространство еще кажется плоским. Но, как известно, предполагаемый избыток или дефект возрастает с увеличением размеров треугольников, и не исключено, что треугольники, размеры которых сравнимы с размерами нашей Галактики (или еще больше), вполне могут обладать заметным избытком или дефектом. Это означало бы, что наш мир искривлен. Существуют и другие косвенные аргументы, подтверждающие гипотезу о кривизне нашего мира. Если эта гипотеза верна, то радиус кривизны Вселенной составляет несколько миллиардов световых лет.

Математически можно показать, что такой мир с положительной кривизной должен быть конечным и замкнутым. Это — замкнутое пространство, подобно тому как сфера — замкнутая поверхность. Все время двигаясь в таком мире по прямой, мы в конце концов вернемся в исходную точку, как и при кругосветных путешествиях на Земле.

Как представить себе замкнутое пространство? Замкнутую поверхность мы можем поместить в трехмерное пространство и рассматривать со всех сторон, но как быть с замкнутым пространством?

Как поступит обитатель сферы, не способный наглядно представить себе шар, желая нарисовать картину своего мира? Он начертит две карты (одну — восточного, другую — западного полушария) и разложит их рядом друг с другом. Заглянув в любой географический атлас, мы без труда обнаружим в нем две такие карты полушарий Земли. Градусная сетка, экватор и другие условные обозначения позволяют понять, каким образом следует отождествлять друг с другом границы карт.

Аналогичным образом можно поступить и с замкнутым трехмерным миром. «Разрежем» его по поверхности какой-нибудь сферы. Наш мир распадется на две части — два шара, заполненных звездами и млечными путями. Затем мы построим модели половинок и поставим их перед собой на письменный стол — два прозрачных глобуса, внутри которых пылинки означают спиральные туманности. Сферическая поверхность, по которой мы разрезали Вселенную, видна на обоих глобусах: это их наружная поверхность. При желании на поверхности глобусов можно нанести координатные сетки и указать, какие точки поверхностей соответствуют друг другу.

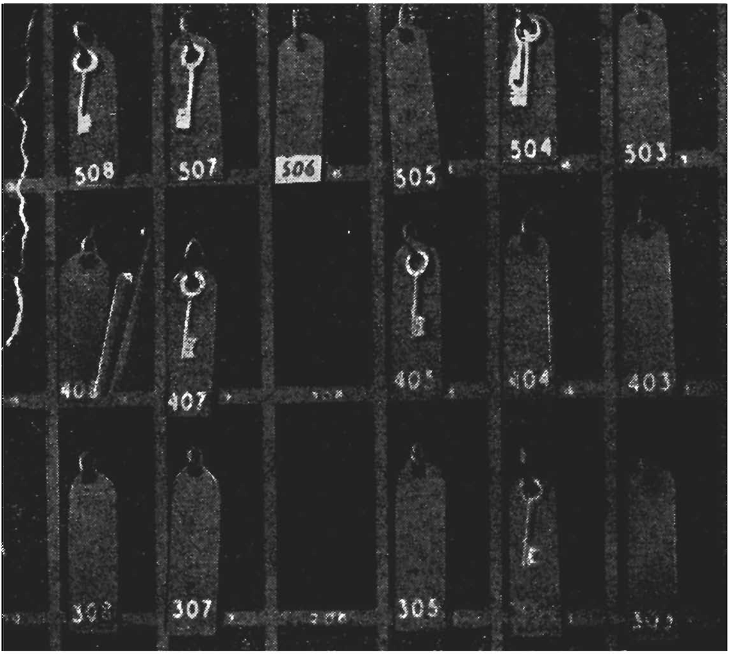
Такой мир действительно замкнут. Каждое путешествие в нем по прямой оказывается кругосветным. Прослеживая кругосветное путешествие по замкнутой Вселенной на нашей модели, мы рано или поздно покинем пределы одного глобуса. Ничего страшного при этом не произойдет, поскольку, добравшись до его границы (наружной поверхности), мы тем самым оказываемся в отождествленной с ней точке, принадлежащей наружной поверхности другого глобуса, а при дальнейшем движении по прямой внедряемся в его глубины. По существу такое происшествие, как пересечение наружной поверхности глобусов в нашей модели замкнутой Вселенной, означает не больше и не меньше, чем переход через экватор или нулевой меридиан при путешествии по Земле. Поверхность самого сечения не является реально существующей границей, отделяющей одну часть Вселенной от другой, границей она выглядит лишь на модели.

С искривленным миром, в который нас поместили астрономы, происходит и еще нечто странное: он раздувается. Расстояния между объектами во Вселенной возрастают, туманные спиральности разбегаются, удаляясь от нас и друг от друга. Все это началось миллиарды лет назад, когда мир был плотно сжатым комком.

Но проследить судьбы нашего мира — удел астрономии, а не геометрии.

Глава 2

До бесконечности



Различные бесконечности

Представим себе гостиницу с бесконечным числом номеров, «перенумерованных» по порядку

$$1, 2, 3, \ldots .$$

Все номера заняты. Поздно вечером приезжает еще один гость. «Свободных мест нет», — говорит ему портье. «Это не играет роли, — вступает в разговор управляющий. — Переселим гостя из номера 1 в номер 2, гостя из номера 2 — в номер 3, гостя из номера 3 — в номер 4 и так далее, а вновь прибывшего гостя поместим в освободившийся номер 1».

Среди ночи приезжает еще 1000 гостей. «Свободных мест нет», — говорит им портье. «Неважно, — возражает управляющий. — Переселим гостя из номера 1 в номер 1001, гостя из номера 2— в номер 1002 и так далее, а вновь прибывших гостей поместим в освободившиеся номера от 1 до 1000».

Не успели все гости разойтись по отведенным им номерам, как в гостиницу вваливается толпа. На этот раз вновь прибывших бесконечно много, и мы обозначим их $A_1, A_2, A_3, \ldots$ «Свободных мест нет», — говорит портье. «Ничего страшного, — все так же спокоен управляющий. — Переселим гостя из номера 1 в номер 2, гостя из номера 2 — в номер 4, гостя из номера 3 — в номер 6 и вообще каждого гостя из последующего номера попросим переехать в номер с вдвое большим числом. Тогда гостей A_1, A_2, A_3, ... мы сможем поселить в номерах 1, 3, 5, ...»

Эту занимательную историю любил рассказывать знаменитый математик Давид Гильберт (1862—1943). Она позволяла ему наглядно объяснить своим слушателям смысл важных понятий, введенных в математику Георгом Кантором (1845—1918).

Последовательность *натуральных чисел*

$$1, 2, 3, \ldots$$

бесконечна: ее можно продолжать неограниченно. Последовательность четных чисел

$$2, 4, 6, \ldots$$

также бесконечна. Четные числа составляют лишь часть всех натуральных чисел, однако их не меньше, чем натуральных чисел (говорят, что мощность четных чисел не меньше мощности натуральных чисел). Действительно, последовательность четных чисел можно расположить под последовательностью натуральных чисел так, что каждому натуральному числу будет соответствовать ровно одно четное число и, наоборот, каждому четному — одно и только одно натуральное число:

$$1,\ 2,\ 3,\ 4,\ \ 5,\ \ldots,$$
$$2,\ 4,\ 6,\ 8,\ 10,\ \ldots.$$

Иначе говоря, всех уже живущих в переполненной до отказа гостинице можно разместить лишь по четным номерам.

Записав в «два этажа» числа

$$\begin{array}{lllll} 1, & 2, & 3, & 4, & \ldots, \\ 1000, & 2000, & 3000, & 4000, & \ldots, \end{array}$$

мы сразу же увидим, что мощность чисел, делящихся на 1000, также равна мощности всех натуральных чисел. То же можно сказать и о натуральных числах, начиная с 1001:

$$\begin{array}{lllll} 1, & 2, & 3, & 4, & \ldots, \\ 1001, & 1002, & 1003, & 1004, & \ldots. \end{array}$$

А как обстоит дело с положительными рациональными числами (то есть с числами, представимыми в виде дробей, например $^7/_3$, $^2/_5$ и так далее)? Запишем положительные рациональные числа под натуральными:

$$\begin{array}{cccccccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & \ldots \\ ^1/_1 & ^1/_2 & ^2/_1 & ^1/_3 & ^3/_1 & ^1/_4 & ^2/_3 & ^3/_2 & ^4/_1 & ^1/_5 & ^5/_1 & \ldots. \end{array}$$

Рациональные числа следуют друг за другом в строго определенном порядке: сначала идут все числа a/b, у которых $a+b=2$, затем числа с $a+b=3$, за ними — числа с $a+b=4$ и так далее, а при фиксированном n числа a/b с $a+b=n$ располагаются в порядке возрастания a. (При этом сократимые дроби следует отбрасывать.)

Итак, мощность рациональных чисел не больше мощности целых чисел.

Бесконечное множество M любых предметов называется *счетным*, если между ним и множеством натуральных чисел можно установить взаимно однозначное соответствие: каждому натуральному числу сопоставить один и только один элемент множества M и, наоборот, каждому элементу множества M сопоставить одно и только одно натуральное число. Иначе говоря, счетным называется множество, элементы которого можно разместить в гостинице с номерами 1, 2, 3, ... так, что каждая комната будет занята одним и только одним элементом множества M.

Все бесконечные множества, о которых мы говорили до сих пор, счетны. Существуют ли другие бесконечные множества? Оказывается, существуют.

Рассмотрим какой-нибудь отрезок прямой A. Он содержит бесконечно много точек. Я утверждаю, что множество точек, образующих отрезок A, не принадлежит к числу тех множеств, которые мы условились называть счетными.

Если бы кто-нибудь вздумал утверждать, что множество точек, образующих отрезок A, вопреки моему утверждению все же счетно, то ему пришлось бы расположить все точки отрезка A, не пропустив ни одной из них, в виде последовательности

$$a_1,\ a_2,\ a_3,\ \ldots.$$

Я покажу, что расположить так все точки отрезка A не удастся ни моему оппоненту, ни любому другому человеку: по крайней мере одна точка, принадлежащая отрезку A, окажется пропущенной.

Итак, мой противник приступает к перечислению точек отрезка A, а пока он этим занимается, я построю точку, которую он непременно пропустит.

Мой противник выбирает какую-то точку отрезка A и обозначает ее a_1. В ответ на это я выбираю *подотрезок* A_1 отрезка A, не содержащий точку a_1 (подотрезком отрезка A мы условимся называть отрезок меньшей длины, все точки которого принадлежат отрезку A). На рис. 24 подотрезок A_1 для удобства изображен отдельно, в действительности же подотрезок A_1 лежит на отрезке A. Затем мой противник выбирает точку a_2 на отрезке A_1, на что я отвечаю выбором подотрезка A_2 отрезка A_1, не содержащим точку a_2. (Подотрезок A_2 для большей ясности изображен на рис. 24 над отрезком A. В действительности он так же, как и подотрезок A_1, лежит на отрезке A.) Когда же мой противник выбирает точку a_3, я выбираю подотрезок A_3 отрезка A_2, не содержащий точку

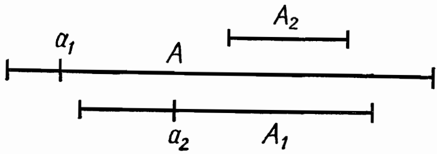


Рис. 24.

a_3. Так мы продолжаем действовать неограниченно долго. При этом каждый раз

A_{n+1} означает подотрезок отрезка A_n,
точка a_n не принадлежит отрезку A_n.

Выбирая свои подотрезки, я слежу за тем, чтобы длина очередного подотрезка A_{n+1} составляла не более половины длины предыдущего подотрезка A_n. Все построенные мной отрезки A_n стягиваются к некоторой точке c. Я утверждаю, что именно эту точку c мой противник непременно должен пропустить.

Действительно, может ли точка c совпадать с точкой a_1? Нет, поскольку точка c принадлежит отрезку A_1, в то время как отрезок A_1 специально был выбран таким образом, чтобы он не содержал точку a_1. Может ли точка c совпадать с точкой a_2? Нет, поскольку точка c принадлежит отрезку A_2, в то время как точка a_2 не принадлежит этому отрезку. Проверку можно было бы продолжать и дальше. Может ли точка c совпадать с точкой a_{1000}? Нет, поскольку точка c принадлежит отрезку A_{1000}, заведомо не содержащему точку a_{1000}. Итак, не ограничивая общности, можно утверждать, что точка c не совпадает ни с одной из точек a_1, a_2, ..., выбранных моим противником. Следовательно, моему противнику не удалось перечислить все точки отрезка A, точку c он все-таки пропустил.

«Я мог бы начать с точки c», — возразит мой противник. (Именно так поступил человек, побившийся об заклад съесть 20 картофелин. Съев 19 из них и чувствуя себя не в силах проглотить последнюю картофелину, этот человек со вздохом заметил: «С нее-то мне и следовало бы начать».)

Но, разумеется, если бы мой противник начал перечисление точек отрезка A с точки c, то я указал бы некую другую точку, также упущенную им из виду. Каким бы способом он ни выбирал счетное множество точек отрезка A, моему противнику все равно не удастся исчерпать весь запас точек, принадлежащих отрезку A.

Мощность множества точек, образующих отрезок A, не меньше мощности счетного множества. В действительности же мощность точек, принадлежащих произвольному отрезку A, больше мощности счетного множества.

Высказанное нами утверждение было впервые высказано Георгом Кантором. Кантор поставил вопрос: «Существуют ли бесконечные множества, мощность которых больше мощности счетных множеств, но меньше мощности множеств точек, заполняющих отрезок прямой?» Самому Кантору так и не удалось найти ответ на этот вопрос. Орешек оказался столь твердым, что на нем сломали зубы многие математики. После упорных, но тщетных усилий решить поставленную Кантором проблему нашлись критически мыслящие люди, которые попытались выяснить, как именно следует ее точно формулировать, и в 1963 г. Пол Коэн доказал, что проблема Кантора неразрешима. Точнее говоря, на поставленный Кантором вопрос можно ответить и утвердительно, и отрицательно, причем ни тот, ни другой ответ не приводит к противоречию. Но этим, насколько можно судить, дело не исчерпывается. По-видимому, мы еще не нашли «правильную» формулировку канторовской проблемы.

Натуральные числа

Вернемся теперь к натуральным числам. Над ними можно производить различные операции, в частности складывать и умножать. Для этого необходимо ввести правила (называемые также аксиомами), которым должны подчиняться арифметические действия над натуральными числами.

Например, мы хотим, чтобы сумма двух натуральных чисел была натуральным числом, в сумме можно было переставлять слагаемые, а в произведении сомножители, не изменяя при этом значения ни суммы, ни произведения. Желательно, чтобы мы могли расставлять и раскрывать скобки. Было бы весьма удобно, если бы существовал единичный элемент относительно умножения, то есть натуральное число, при умножении на которое любое другое натуральное число сохраняло бы свое значение.

Мы рассмотрим лишь некоторые из аксиом арифметики натуральных чисел и прежде всего хорошо известный нам из практики *коммутативный (или переместительный) закон сложения*

$$a+b=b+a.$$

Для всех ли натуральных чисел он выполняется? Что можно сказать об *ассоциативном (переместительном) законе сложения*

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

позволяющем группировать при вычислениях слагаемые наиболее удобным способом? Можно ли доказать его? Ведь за краткой — в одну строку — формулой

$$(a+b)+c=a+(b+c)$$

кроется бесконечно много соотношений между конкретными натуральными числами.

Каким образом можно доказать правильность бесконечного множества соотношений?

Прежде чем приступить к поиску ответов на эти вопросы, необходимо точно определить, в каком смысле следует понимать выражение $a+b$. Складывая два натуральных числа, мы получаем новое натуральное число, но какое? Чтобы ответить на этот вопрос, мы должны были бы перебрать все возможные пары натуральных чисел и для каждой пары вычислить сумму образующих ее чисел. Каким образом это можно сделать?

Ряд натуральных чисел начинается с 1. Для каждого члена ряда натуральных чисел существует один и только один следующий за ним член того же ряда, а для всех членов натурального ряда, кроме 1, — один и только один предыдущий член. Пусть n' — число, следующее за n.

Теперь мы уже достаточно подготовлены для того, чтобы ввести точное определение суммы двух натуральных чисел $a+b$:

$$a+b=a' \quad \text{при} \quad b=1,$$
$$a+b=(a+n)' \quad \text{при} \quad b=n'.$$

Это означает, что число $a+1$ следует за числом a. Если число $a+n$ уже известно, то следующее за ним число равно сумме $a+$ число, следующее за числом n. Иначе говоря,

$$a+(n+1)=(a+n)+1. \qquad (K_1)$$

Полученное соотношение выражает ассоциативный закон сложения в частном случае: при $b=n$ и $c=1$.

Докажем теперь ассоциативный закон сложения для общего случая:

$$a+(b+c)=(a+b)+c. \qquad (K_c)$$

Соотношение K_1 мы доказали ранее. Следовательно, при $c=1$ соотношение K_c выполняется.

Предположим, что оно верно при $c=n$ и любых a, b, то есть что выполняется соотношение

$$a+(b+n)=(a+b)+n. \qquad (K_n)$$

Воспользуемся теперь соотношением K_n для доказательства нового соотношения

$$a+(b+n')=(a+b)+n'. \qquad (K_{n'})$$

Для удобства условимся после каждого преобразования указывать, на основании чего оно выполнено.

$$\begin{aligned}
a+(b+n')\,a &= a+(b+(n+1))= && \text{(по определению } n'\text{)}\\
&= a+((b+n)+1)= && (K_1)\\
&= (a+(b+n))+1= && (K_1)\\
&= ((a+b)+n)+1= && (K_n)\\
&= (a+b)+(n+1)= && (K_1)\\
&= (a+b)+n'. && \text{(по определению } n'\text{)}
\end{aligned}$$

Каков же итог? Мы начали с доказательства соотношения K_1, затем предположили, что выполняется соотношение K_n, и, используя это предположение, доказали соотношение K_n'. Отсюда следует, что соотношение K_c выполняется при любом c. Тем самым ассоциативный закон сложения натуральных чисел

$$a + (b + c) = (a + b) + c \qquad (K_c)$$

доказан.

Избранный нами способ доказательства называется *доказательством по индукции*, или заключением от n к $n+1$: известно, что соотношение K_1 выполнено, мы предполагаем, что выполняется соотношение K_n, и доказываем, что если наше предположение верно, то выполняется также и соотношение K_{n+1}. Таким образом, соотношение K_n выполняется при любом значении индекса n: из соотношения K_1 следует соотношение K_2, из K_2 — соотношение K_3 и так далее до бесконечности.

Обобщая, мы могли бы сформулировать *принцип полной математической индукции.*

Известно, что утверждение E_1 истинно.

Предположим, что утверждение E_n истинно.

Докажем, что если это предположение верно, то утверждение E_{n+1} также истинно.

Тогда утверждение E_n истинно при любых значениях индекса n.

Воспользуемся теперь принципом полной математической индукции для доказательства *коммутативного закона сложения* для натуральных чисел.

Начнем с рассмотрения частного случая: докажем соотношение

$$a + 1 = 1 + a. \qquad (L_a)$$

При $a = 1$ утверждение очевидно. Предположим, что выполняется соотношение

$$n + 1 = n + 1. \qquad (L_n)$$

Тогда

$$(n + 1) + 1 = (1 + n) + 1 = \qquad (L_n)$$

$$= 1 + (n + 1). \qquad (K_1)$$

Следовательно, соотношение L_{n+1} также выполняется, что и требовалось доказать. Иначе говоря, проверив (оно оказалось самоочевидным) соотношение L_1 и предположив, что выполняется соотношение L_n, мы доказали соотношение L_{n+1}. Тем самым соотношение L_a доказано для любого натурального значения индекса a. Докажем теперь коммутативный закон сложения натуральных чисел для общего случая

$$a+b=b+a. \qquad (M_b)$$

Соотношение M_1 совпадает с только что доказанным соотношением L_a. Предположим, что выполняется соотношение

$$a+n=n+a. \qquad (M_n)$$

Тогда

$$a+(n+1)=(a+n)+1=(n+a)+1=$$
$$=1+(n+a)=(1+n)+a=(n+1)+a.$$

Итак, соотношение $M_1 \equiv L_a$ верно, и, предположив, что выполняется соотношение M_n, мы доказали соотношение M_{n+1}. Тем самым доказано, что соотношение M_b выполняется при любых натуральных значениях индекса b.

Теперь можно ввести умножение натуральных чисел, определить число 1 как единичный элемент относительно операции умножения и воспользоваться теми следствиями, к которым приводят новые понятия.

Определение.

$$a\cdot 1=a,$$
$$a(n+1)=an+a.$$

Пользуясь принципом полной математической индукции, нетрудно показать, что произведение ab тем самым определено для любых натуральных чисел a и b.

Докажем некоторые свойства введенной нами операции умножения.

Начнем с доказательства правила

$$(a+b)c=ac+bc. \qquad (N_c)$$

Соотношение N_1 выполняется, поскольку $a\cdot 1=a$, $b\cdot 1=b$.

Предположим, что выполняется соотношение

$$(a+b)n=an+bn. \qquad (N_n)$$

Тогда

$$(a+b)(n+1)=(a+b)n+(a+b)=\text{(по определению произведения)}$$
$$=(an+bn)+(a+b).$$

Повторно применив коммутативный и ассоциативный законы, последнее выражение можно привести к виду

$$(an+a)+(bn+b).$$

По определению произведения двух натуральных чисел это не что иное, как

$$a(n+1)+b(n+1).$$

Итак, соотношение N_{n+1} доказано, если выполняется соотношение N_n. Тем самым соотношение N_c можно считать доказанным при любых натуральных значениях индекса c.

Докажем, что

$$1\cdot a=a. \qquad (P_a)$$

Соотношение P_1 выполняется по определению произведения двух натуральных чисел. Из соотношения

$$1\cdot n=n \qquad (P_n)$$

следует, что

$$1\cdot(n+1)=1\cdot n+1=n+1.$$

Тем самым соотношение P_{n+1}, а вместе с ним и соотношение P_a при любом натуральном значении индекса a доказано.

Наконец, докажем коммутативный закон умножения натуральных чисел:

$$ab=ba \qquad (Q_b)$$

Прежде всего заметим, что $Q_1=P_a$. Из соотношения

$$an=na \qquad (Q_n)$$

следует, что

$$a(n+1)=an+a=na+a=na+1\cdot a.$$

Последнее выражение в силу доказанного выше соотношения N_a, записанного для случая $a=n$ и $b=1$, равно

$$(n+1)a.$$

Таким образом, соотношение Q_{n+1} следует из соотношения Q_n. Тем самым соотношение Q_b доказано при любых натуральных значениях индекса b.

Целые числа

Итак, на вопрос, каким образом можно доказать правильность бесконечного множества соотношений, мы ответили: воспользоваться методом полной математической индукции.

Этот же метод можно с успехом применять и в тех случаях, когда нам нужно определить бесконечное множество каких-то элементов. Чтобы определить a_n, мы полагаем

$$a_1 = a \quad \text{и} \quad a_{n+1} = a_n a.$$

Тем самым a_n определено при всех натуральных значениях n. Ранее мы аналогичным образом определили сложение:

$$a + 1 = a' \quad \text{и} \quad a + n' = (a + n)'.$$

Мы установили правила, которые необходимо соблюдать при выполнении операции сложения (ввели аксиомы сложения), и вывели из них другие арифметические правила.

Операция, обратная сложению (вычитание), не определена в множестве натуральных чисел, поскольку это множество не содержит отрицательных чисел. Для того чтобы мы могли определить разность любых двух натуральных чисел, множество натуральных чисел необходимо расширить, дополнив его новыми элементами. Каждому натуральному числу n мы должны сопоставить новое число $-n$ и, присоединив все новые числа к множеству натуральных чисел, дополнить его еще одним элементом — числом 0. Получившееся множество чисел

$$\ldots, -2, -1, 0, 1, 2, \ldots$$

называется множеством *целых чисел.*

На целые числа без труда переносятся уже известные нам коммутативный и ассоциативный законы сложения натуральных чисел. Для целых чисел определены операции умножения и вычитания. В качестве единичного элемента относительно сложения можно определить лишь число 0 так, чтобы

$$a+0=a,$$

а элемент $-a$, обратный числу a, так, чтобы

$$a-a=0.$$

Мы уже располагаем всем необходимым для того, чтобы ввести определение абсолютной величины $|a|$ целого числа a: абсолютная величина неотрицательного числа a равна самому числу, а абсолютная величина отрицательного числа a равна обратному числу $-a$, то есть

$$|a|=a \quad \text{при} \quad a \geqslant 0,$$
$$|a|=-a \quad \text{при} \quad a<0.$$

Какие аксиомы мы ввели для того, чтобы над целыми и натуральными числами можно было производить операции сложения и умножения? Сложив или умножив целые числа, мы снова получим целое число. Для сложения и умножения целых чисел выполняются ассоциативный и коммутативный законы. Относительно операции сложения мы определили единичный элемент (число 0) и ввели понятие обратного элемента, а относительно операции умножения ввели единичный элемент (число 1). Но можем ли мы указать элемент, обратный данному целому числу относительно умножения, то есть определить над целыми числами операцию *деления*? Нет, поскольку лишь в редких случаях при делении одного целого числа на другое мы получим целое число. Если мы хотим ввести деление, то множество целых чисел необходимо расширить. Тут-то и появляются рациональные числа, с которыми мы уже встречались в начале этой главы. Множество рациональных чисел состоит из всех целых чисел и чисел, представимых в виде отношения двух целых чисел. Число 1 мы определили как единичный элемент относительно операции умножения, удовлетворяющий соотношению

$$a \cdot 1=a.$$

Рациональное число $1/a$ называется обратным числу a относительно умножения, если

$$a \cdot \frac{1}{a}=1.$$

Итак, мы ввели основные правила арифметики, но пока еще не знаем, достаточны ли они, то есть можно ли считать, что перечисленные нами аксиомы однозначно определяют сложение и вычитание, умножение и деление для любых a и b. Доказательство достаточности наших правил по сложности намного превосходит любое из приведенных выше доказательств, и мне не хотелось бы обременять читателя излишними подробностями.

Простые числа

Числа

$$2, 3, 5, 7, 11, 13, 17, 19, 23, 29, 31, \ldots$$

в ряду натуральных чисел называются *простыми*, поскольку их нельзя представить в виде произведения других натуральных чисел. Такие числа, как

$$4 = 2 \cdot 2, \quad 6 = 2 \cdot 3, \quad 8 = 2 \cdot 2 \cdot 2, \quad 9 = 3 \cdot 3, \ldots$$

не являются простыми. Если натуральное число n не простое, то его можно разложить на множители и, продолжая разлагать в случае необходимости каждый множитель, представить число n в виде произведения простых чисел. Число 1 не считается простым.

Сколько существует простых чисел? Обрывается ли где-нибудь ряд простых чисел? Один из наиболее часто применяемых в математике способов доказательства состоит в том, что вместо прямого утверждения пытаются доказать противоположное ему утверждение. Если такая попытка наталкивается на противоречие, то отсюда, как учит нас логика, можно заключить, что прямое утверждение верно.

Именно этим способом доказательства мы сейчас и воспользуемся.

Предположим, что ряд простых чисел обрывается. Тогда существует лишь конечно много простых чисел

$p_1, p_2, \ldots, p_k$, где k — некое конечное число.

Составим произведение всех простых чисел и прибавим к нему 1:

$$P = p_1 p_2 \ldots p_k + 1.$$

Число P больше любого из простых чисел $p_1, p_2, \ldots$ $\ldots, p_k$. Поскольку других простых чисел, кроме p_1,

$p_2, \ldots, p_k$, не существует, то P не может быть простым числом, и, следовательно, его можно представить в виде произведения простых чисел. Заведомо можно утверждать, что какое-то простое число входит в разложение числа P. Поскольку числами $p_1, p_2, \ldots$ $\ldots, p_k$ по предположению исчерпываются все простые числа, то число P должно делиться на одно из чисел $p_1, p_2, \ldots, p_k$. Но это невозможно: при делении числа P на любое из чисел $p_1, p_2, \ldots, p_k$ мы получим остаток, равный 1. Следовательно, наше исходное предположение о том, что существует лишь конечное множество простых чисел, ошибочно.

Итак,

существует бесконечно много простых чисел.

Эта теорема была известна еще Евклиду (IV в. до н. э.).

Рассмотрим числа вида $4n-1$, где $n = 1, 2, 3, \ldots$:

$$3,\ 7,\ 11,\ 15,\ 19,\ 23,\ 27,\ 31,\ \ldots.$$

Можно ли утверждать, что среди них существует бесконечно много простых чисел?

Числа вида $4n-1$ нечетны. Остальные нечетные числа представимы в виде $4n-3$. Умножив друг на друга два числа вида $4n-3$, мы снова получим число того же вида:

$$(4k-3)(4m-3) = 4(4km-3k-3m+3)-3,$$

поскольку выражение, стоящее в круглых скобках правой части, при натуральных k и m принимает лишь натуральные значения. Следовательно, число вида $4n-1$ нельзя разложить лишь на множители вида $4n-3$: по крайней мере один из множителей должен быть числом вида $4n-1$. Разлагая число $4n-1$ на простые множители, мы также должны получить по крайней мере одно простое число вида $4n-1$. Таким образом, число вида $4n-1$ должно иметь по крайней мере один простой делитель, также имеющий вид $4n-1$.

Умножив два числа вида $4n-1$ одно на другое, мы получим

$$(4k-1)(4m-1) = 4(4km-4k-4m+1)-3,$$

то есть число вида $4n-3$. Следовательно, произведение четного числа сомножителей вида $4n-1$ имеет вид $4n-3$. Прибавив к такому произведению 2, мы получим число вида $4n-1$.

Предположим, что существует лишь конечно много простых чисел вида $4n-1$. Составим полный список таких чисел:

$$p_1, p_2, \ldots, p_k.$$

Поскольку мне очень хочется, чтобы индекс k был четным, я в случае необходимости могу включить число 3 в список простых чисел вида $4n-1$ дважды, то есть положить $p_1=p_2=3$, $p_3=7$ и так далее. Число

$$P=p_1p_2 \ldots p_k+2,$$

как мы уже знаем, имеет вид $4n-1$. Это число не может быть простым, так как оно больше любого из простых чисел $p_1, p_2, \ldots, p_k$. Поскольку число P представимо в виде $4n-1$, то в его разложение на простые множители должно входить одно простое число вида $4n-1$, встречающееся среди чисел p_1, $p_2, \ldots, p_k$. Но это невозможно, поскольку при делении числа P на любое из чисел $p_1, p_2, \ldots, p_k$ мы получим остаток 2. Следовательно, исходное предположение неверно. Итак,

существует бесконечно много простых чисел вида $4n-1$.

Простых чисел вида $4n-3$ также бесконечно много, но доказать это гораздо труднее. Справедливо и более общее утверждение:

если начальный член a и разность d арифметической прогрессии

$$a, a+d, a+2d, a+3d, \ldots$$

не имеют общих делителей, то среди ее членов содержится бесконечно много простых чисел (число 1 не считается простым).

Доказать эту теорему очень трудно.

Рациональные и иррациональные числа

Взяв отрезок прямой, мы можем отложить отрезок вдвое, втрое и вообще в n раз большей длины (где n — любое натуральное число), разделить его пополам, на три части и на k частей (где k — любое натуральное число). Начав с единичного отрезка, нетрудно построить отрезок длины m/n, где m и n — натуральные числа. Такие отрезки соответствуют рациональным числам, которые, как мы уже знаем, образуют счетное множество. Но известно также, что множество точек, заполняющих любой отрезок прямой, обладает большей мощностью. Следовательно, существуют и такие отрезки, длину которых нельзя выразить рациональным числом. Они соответствуют так называемым иррациональным числам. Множество иррациональных чисел дополняет множество рациональных чисел. Взятые вместе, эти множества образуют множество *действительных чисел.*

Построить отрезок, длина которого выражается иррациональным числом, совсем нетрудно. Пользуясь теоремой Пифагора, можно вычислить, что диагональ квадрата со стороной 1 равна $\sqrt{2}$. Число $\sqrt{2}$ нерациональное. Действительно, если бы оно было рациональным, то есть представимым в виде

$$\sqrt{2}=\frac{p}{q} \qquad (*)$$

с взаимно простыми (то есть не имеющими отличного от 1 общего делителя) числами p и q, то, возводя в в квадрат обе части равенства, мы получили бы

$$2=\frac{p^2}{q^2},$$

или

$$2q^2=p^2.$$

Следовательно, число p^2, а значит и q^2, четно, то есть представимо в виде

$$p=2p',$$

где p' — некоторое натуральное число.

Но тогда

$$2q^2=4p'^2,$$

$$q^2=2p'^2.$$

Таким образом, q^2 и, следовательно, q — также четные числа.

Итак, вопреки исходному предположению о том, что числа p и q взаимно простые, они оба оказались четными. Иначе говоря, исходное предположение приводит к противоречию. Следовательно,

$\sqrt{2}$ — *иррациональное число.*

Но нам хотелось бы знать больше: какова длина диагонали единичного квадрата? Нельзя ли вычислить $\sqrt{2}$?

По определению число $\sqrt{2}$ является положительным решением уравнения

$$x^2 = 2.$$

Хотя это уравнение не допускает точных решений в рациональных числах, ничто не мешает нам искать его приближенное решение. При умелом подходе можно довольно быстро найти весьма хорошее приближение.

Итак, нам нужно построить квадрат, площадь которого в выбранных единицах измерения равнялась бы 2. Построим вместо квадрата прямоугольник со сторонами, незначительно отличающимися по длине (длины обеих сторон выражаются рациональными числами).

Если длину одной стороны прямоугольника выбрать

$$a_1 = 1,$$

то длина другой стороны должна быть

$$b_1 = 2.$$

Это еще плохое приближение: длины сторон заметно отличаются друг от друга. Выберем одну сторону нового прямоугольника равной полусумме сторон построенного прямоугольника, то есть положим

$$a_2 = \frac{1}{2}(a_1 + b_1) = \frac{3}{2},$$

тогда

$$b_2 = \frac{2}{a_2} = \frac{4}{3}.$$

Площадь прямоугольника со сторонами $3/2$ и $4/3$ равна 2, и по форме он уже больше напоминает квадрат. Не довольствуясь достигнутым, продолжим наши поиски. Построим еще один прямоугольник со сторонами

$$a_3 = \frac{1}{2}(a_2 + b_2) = \frac{17}{12},$$

$$b_3 = \frac{2}{a_3} = \frac{24}{17}.$$

Его площадь, так же как и площадь предыдущего прямоугольника, равна 2, а длины сторон отличаются лишь на $1/204$. Следующий прямоугольник будет иметь уже стороны

$$a_4 = \frac{1}{2}(a_3 + b_3) = \frac{577}{408},$$

$$b_4 = \frac{2}{a_4} = \frac{816}{577}.$$

Их длины отличаются на $1/235416$. Построение прямоугольников можно продолжать сколь угодно долго. Стороны $(n+1)$-го прямоугольника связаны со сторонами n-го прямоугольника соотношениями

$$a_{n+1} = \frac{1}{2}(a_n + b_n),$$

$$b_{n+1} = \frac{2}{a_{n+1}}.$$

Такой способ определения членов числовой последовательности называется *рекуррентным.*

Длины сторон n-го прямоугольника выражаются рациональными числами:

$$a_n = \frac{p_n}{q_n}, \quad b_n = \frac{2q_n}{p_n}.$$

Длину стороны a_{n+1} $(n+1)$-го прямоугольника можно представить в виде

$$a_{n+1} = \frac{p_n^2 + 2q_n^2}{2p_nq_n}.$$

Следовательно,

$$p_{n+1} = p_n^2 + 2q_n^2,$$

$$q_{n+1} = 2p_nq_n.$$

Оценим, насколько отличаются по длине стороны n-го прямоугольника, то есть разность

$$a_n - b_n = \frac{p_n^2 - 2q_n^2}{p_n q_n}.$$

Рассмотрим сначала числитель дроби, стоящей в правой части

$$z_n = p_n^2 - 2q_n^2.$$

Величину z_{n+1} можно записать в виде

$$z_{n+1} = p_{n+1}^2 - 2q_{n+1}^2 = (p_n^2 + 2q_n^2)^2 - 8p^2q^2 = (p_n^2 - 2q_n^2)^2.$$

Следовательно,

$$z_{n+1} = z_n^2.$$

Но $a_1 = 1$, поэтому $p_1 = q_1 = 1$ и $z_1 = -1$. При последующих возведениях в квадрат мы получаем при всех $n > 1$

$$z_n = 1.$$

Следовательно,

$$a_n - b_n = \frac{1}{p_n q_n}.$$

Разность сторон быстро убывает. Чтобы убедиться в этом, оценим знаменатель дроби, стоящей в правой части. Величина q_n удовлетворяет неравенству

$$q_n \geqslant 2^{2^{n-1}-1}.$$

Воспользуемся для доказательства этого неравенства методом полной математической индукции. При $n = 1$ неравенство выполняется. Предположим, что оно выполняется при некотором n. Тогда, поскольку $p_n \leqslant 1$,

$$q_{n+1} = 2p_n q_n > 2q_n^2 = 2 \cdot \left(2^{2^{n-1}-1}\right) = 2^{2^n-1}.$$

Неравенство доказано. Но если

$$q_n \geqslant 2^{2^{n-1}-1},$$

то знаменатель $p_n q_n$ дроби $a_n - b_n$ удовлетворяет неравенству

$$p_n q_n \geqslant q_n^2 \geqslant 2^{2^n-2},$$

а сама разность — неравенству

$$a_n - b_n = \frac{1}{p_n q_n} \leqslant 2^{-2^n+4}.$$

Отсюда видно, что с возрастанием n она действительно быстро убывает.

Длины сторон прямоугольников мы выбирали так, чтобы

$$a_n b_n = 2.$$

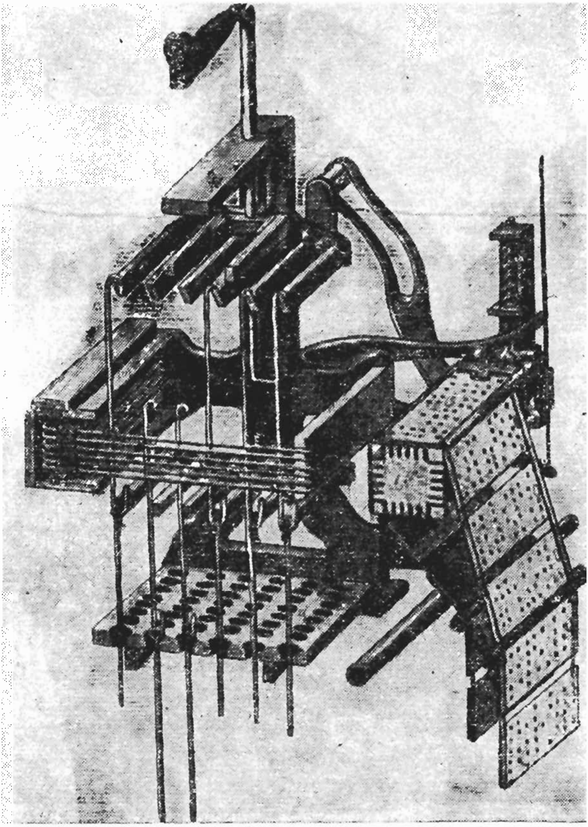
При этом $a_n (n \geqslant 1)$ несколько больше, а b_n — несколько меньше $\sqrt{2}$. Число $\sqrt{2}$ расположено между числами a_n и b_n, которые при достаточно большом n сколь угодно мало отличаются одно от другого. Таким образом, число $\sqrt{2}$ можно заключить между двумя сколь угодно близкими рациональными числами, но само оно не рациональное. Рациональные приближения к числу $\sqrt{2}$ можно строить неограниченно долго — до бесконечности.

Любое действительное число также можно заключить между двумя рациональными числами и, зажав между верхней и нижней границей, определить сколь угодно точно, но тем не менее приближенно.

Тем самым мы охватили всю числовую прямую, простирающуюся в обе стороны до бесконечности.

Глава 3

Чем занимаются вычислительные машины



Системы счисления

Посмотрите на рис. 25, вы видите на нем, как записали бы число 1967 древний египтянин, житель Вавилона, грек гомеровских времен, грек времен Перикла, европеец эпохи Карла Великого и наш современник. Египтянин расположил бы справа налево знаки, обозначающие число тысяч, сотен, десятков и

Рис. 25.

единиц. Грек времен Гомера (четвертая строка сверху) расположил бы слева направо знаки, указывающие тысячи, полутысячи, сотни, полусотни, десятки, пятерки и единицы (то есть число 1000, 500, 100, 50, 10, 5, 1). Современник Перикла (пятая строка сверху) записал бы число 1967 экономнее, но для этого ему понадобился бы весь греческий алфавит (чтобы обозначать числа 1, 2, ..., 10, 20, ... , 100, 200, ..., 900), и, лишь дойдя до 1000, он вновь вернулся бы к букве α (ранее обозначавшей 1), отметив ее на этот раз штрихом. Так называемые римские цифры, которыми число 1967 записано в шестой строке, вряд ли нуждаются в пояснениях, а индо-арабская запись в последней строке знакома всем нам.

Для нас наибольший интерес представляет вавилонская запись числа 1967, стоящая в третьей строке. Семь клинописных знаков справа в виде треугольников, обращенных основанием вверх, соответствуют семи единицам. Четыре примыкающих к ним слева угла с отчеркнутой вершиной означают четыре десятка, а каждый из двух следующих за ними клинописных знаков, расположенных один под другим, соответствует 60. Наконец, каждый из трех углов с отчеркнутой вершиной, изображенных слева, означает 600. Сложив все вместе, мы получим $3\cdot600+2\cdot60+$ $+4\cdot10+7\cdot1=1967$.

Вавилоняне использовали шестидесятичную систему не только для записи чисел, но и в своей денежной системе, мерах и весах. «Следы» их пристрастия к числу 60 сохранились и поныне. То, что мы делим часы и угловые градусы на 60 минут, а минуты — на 60 секунд, — наследие Древнего Вавилона.

Разумеется, сам по себе факт, что число 60 играло у вавилонян столь большую роль, заслуживает внимания. Но еще более замечательно, что для обозначения $1, 60, 60^2, \ldots$, а также $60^{-1}, 60^{-2}, \ldots$ и $10, 10\cdot$ $\cdot60, \ldots$ вавилоняне использовали одни и те же символы. Примеры такого рода обозначений хорошо известны и нам самим: в числе 66 одна цифра 6 означает 6, а другая — 60.

Наша система счисления называется *позиционной*. Одна и та же цифра, сдвинутая на один знак влево, увеличивает свое значение в 10 раз, а сдвинутая на один знак вправо (при сдвиге вправо в случае необходимости вводится десятичная запятая) — составляет лишь $^1/_{10}$ от своего предыдущего значения. Система счисления, которую применяли вавилоняне, также была позиционной, но ее основанием вместо привычного нам числа 10 служило число 60. Кроме того, древние вавилоняне не знали нуля, что не всегда оказывалось удобным.

С чем связано появление числа 10 в названии нашей системы счисления? Разумеется, с тем, что в нашем языке названия чисел основаны на десятичной системе, хотя сами по себе отнюдь не образуют позиционной системы: для обозначения чисел 1, 10, 100, 1000 и 1 000 000 мы используем слова, не имеющие между собой ничего общего. В свою очередь «десятичная» система названий, придуманных нашими предками для обозначения чисел, появилась по той простой причине, что люди учились считать по пальцам рук и ног. С чисто математической точки зрения наша десятичная система счисления столь же произвольна и условна, как и шестидесятеричная система вавилонян.

Двоичная система

С математической точки зрения наиболее естественной была бы двоичная система счисления, оперирующая лишь двумя цифрами: 0 и 1. Как выглядело бы число 1967 в двоичной системе? На первый взгляд несколько необычно:

$$11110101111.$$

Проверим! Первая единица справа означает (десятичное) число 1, вторая — число 2, третья — 4, четвертая — 8 и так далее. Самая левая единица означает $2^{10} = 1024$. «Перевод» десятичного числа 1967 — задача весьма несложная:

1967	
983	1
491	1
245	1
122	1
61	0
30	1
15	0
7	1
3	1
1	1
0	1

Каждый раз мы делим очередное число на 2 и записываем рядом остаток (равный 0 или 1). Если все выписанные остатки записать не сверху вниз, а слева направо, то получится двоичный перевод исходного десятичного числа.

Двоичная система чрезвычайно удобна на практике. Таблицы сложения и умножения выглядят в двоичной системе так:

$$0+0=0, \quad 0+1=1, \quad 1+0=1, \quad 1+1=10,$$
$$0\cdot 0=0, \quad 0\cdot 1=0, \quad 1\cdot 0=0, \quad 1\cdot 1=1,$$

Восемь равенств вместо двухсот, которые приходится заучивать наизусть в десятичной системе!

Всякий, кто знает двоичную систему, может производить в ней действия так же, как в десятичной:

$$+\begin{array}{r} 100101 \\ 11110 \\ \hline 1000011 \end{array} \qquad -\begin{array}{r} 1000011 \\ 11110 \\ \hline 100101 \end{array} \qquad \times\begin{array}{r} 1101 \\ 101 \\ \hline 1101 \\ 1101 \\ \hline 1000001 \end{array}$$

(Здесь решены следующие примеры на сложение, вычитание и умножение: $37+30=67$, $67-30=37$, $13\cdot 5=65$.) Никаких трудностей не возникает в двоичной системе и с делением, что выгодно отличает ее от десятичной системы. Необходимо лишь каждый раз проверять, можно ли вычесть делитель или нельзя, и записывать в частное в первом случае 1, а во втором — 0. Например,

$$\begin{array}{l|l} 1000001 & 101 \\ \cline{2-2} \underline{101} & 1101 \\ 110 & \\ \underline{101} & \\ 101 & \\ \underline{101} & \end{array}$$

Ним

Небольшой экскурс в сторону от основной темы позволит нам продемонстрировать возможности, таящиеся в двоичной системе.

Есть такая игра — ним, которая пришла к нам, насколько можно судить, из Китая. Играют в ним следующим образом. На столе несколькими кучками раскладывают спички. Двое игроков по очереди делают ходы. Тот, чья очередь делать ход, забирает спички из какой-нибудь *одной* кучки (он должен взять по крайней мере одну спичку, но может забрать и все спички, лежащие в *одной* кучке). Выигрывает тот, кто забирает последнюю спичку.

Предположим, что спички разложены по четырем кучкам: в первой — 10 спичек, во второй — 4, в третьей — 7 и в четвертой — 3 спички. Я делаю первый ход и забираю все 10 спичек из первой кучки. Какой ответный ход предпримете вы? Предположим, что вы забираете 2 спички из третьей кучки и на столе остаются одна кучка из 4, другая — из 5 и третья — из 3 спичек. Я забираю 1 спичку из последней кучки с 3 спичками. Ваш ход? Вы забираете 2 спички из кучки с 5 спичками. На столе остаются одна кучка с 4 спичками, другая — с 3 спичками и одна «кучка», состоящая из одной-единственной спички. Очередным ходом я забираю 2 спички из кучки с 4 спичками. А вы? Вы сдаетесь.

В чем хитрость игры в ним? Запишем данные числа 10, 4, 7, 3 в двоичной системе:

$$\begin{array}{r} 1010 \\ 100 \\ 111 \\ 11 \end{array}$$

В первом столбце слева стоит лишь одна единица, во втором — две единицы, в третьем три единицы и в четвертом — снова две единицы. Мы вправе утверждать, что по крайней мере в одном столбце число единиц нечетно. Всякий раз, когда двоичная запись начального расклада спичек по кучкам содержит по крайней мере один столбец с нечетным числом единиц, выигрывает тот, кто делает первый ход. Ему необходимо лишь следить за тем, чтобы после его хода число единиц во всех столбцах становилось четным. Сделав свой первый ход, я позаботился об этом и забрал все спички из первой кучки. Разумеется, вы уже догадались, какой тактики выгоднее всего придерживаться вам: после очередного вашего хода каждый раз должен возникать столбец с нечетным числом единиц.

После вашего хода единицы в двоичной записи числа оставшихся спичек распределились так:

$$\begin{array}{r} 1\ 0\ 0 \\ 1\ 0\ 1 \\ 1\ 1 \end{array}$$

Во втором столбце число единиц нечетно. Я забираю две спички и тем самым «стираю» лишнюю единицу во втором столбце: теперь во всех столбцах число единиц четно. Ответным ходом вы создаете следующее распределение единиц:

$$\begin{array}{r} 1\ 0\ 0 \\ 1\ 1 \\ 1 \end{array}$$

Так, поочередно делая ходы, мы то создаем столбец с нечетным числом единиц, то уничтожаем его. Игра заканчивается, когда в каждом столбце остается наименьшее четное число единиц, а именно 0 единиц. Именно этого мне и удалось добиться.

Если бы игра начиналась с трех кучек, содержащих

$$\begin{array}{r} 27 = 11011, \\ 5 = \quad 101, \\ 17 = 10001 \end{array}$$

спичек, то первым ходом я забрал бы

$$20 = 10100$$

спичек из первой кучки, и, какие бы ходы вы ни делали, мой выигрыш обеспечен, если только я не совершу какой-нибудь ошибки. Наоборот, если игра начинается с трех кучек, содержащих

$$\begin{array}{r} 31 = 11111, \\ 7 = \quad 111, \\ 24 = 11000 \end{array}$$

спичек, то ни я, ни кто-нибудь другой, делающий первый ход, выиграть не смогут, если только вы не допустите ошибки.

Быстрее, еще быстрее!

Сколько будет 7489·9126? Отвечайте быстро, еще быстрее! Возьмем секундомер и засечем, сколько времени уйдет у нас на выкладки. Мы можем по праву гордиться, если окажется, что произведение двух четырехзначных чисел нам удалось найти не более чем за полминуты: ведь для этого необходимо выписать 27 цифр!

А сколько времени потребовалось бы нам, чтобы решить миллион таких задачек на умножение? В одном году — около полумиллиона минут. Следовательно, выкладки заняли бы у нас целый год — год, в течение которого мы не могли бы ни есть, ни пить, ни спать, ни даже почувствовать себя усталыми (на это у нас просто не оставалось бы времени!). Как, по-вашему, много бы результатов при проверке оказались бы верными?

Современная электронная вычислительная машина решает за полминуты около 10 000 таких задачек, а за час — более миллиона, и это не предел: каждое новое поколение машин считает быстрее предыдущего. В весьма недалеком будущем машины смогут решать за полминуты миллион таких «примеров на умножение».

Возникает вопрос: а нужно ли нам так торопиться?

Не могли бы вы немного подождать?

Производить вычисления — занятие увлекательное, но утомительное, которое требует изрядных затрат времени. Человеку свойственно ошибаться, и когда люди считают, они довольно часто это делают. А получить неверный результат в большинстве случаев гораздо хуже, чем не иметь результата вообще.

С незапамятных времен для вычислений пользовались различными приспособлениями и приборами, облегчающими счет. Вычислять на бумаге люди стали лишь сравнительно недавно. В древности и в средние века вычисления производили, передвигая фишки на счетных досках или перекладывая шарики по клеткам абака. В Китае и поныне пользуются своеобразным «счетным прибором» — китайскими счетами. Отдельные механические счетные устройства появились в

XVII в. В XIX в. началось фабричное производство механических счетных машин. В XX в. их место заняли электромеханические и электронные вычислительные машины.

Но математика и ее приложения развивались гораздо быстрее, чем вычислительные машины. Производить вычисления вовсе не означает заниматься математикой. Сошлюсь хотя бы на то, что в этой книге мы довольно много говорим о математике, но почти не занимаемся выкладками.

Рассмотрим несколько примеров.

В своем великом труде «Математические начала натуральной философии» Ньютон установил законы движения, на которых и поныне зиждется небесная механика. Эти законы можно облечь в формулы, записать в виде так называемых дифференциальных уравнений, а затем в тех случаях, когда требуется вычислить движение небесных тел, решать эти уравнения. Из астрономических наблюдений известны положения и скорости тел нашей Солнечной системы в определенный момент времени, и тот, кто их знает, может вычислить, где находились и где будут находиться в любой момент времени планеты, Луна, кометы. Сам Ньютон мог лишь в общих чертах убедиться в том, что его теория согласуется с наблюдениями. Ни создатель «Математических начал», ни его последователи не ставили перед собой задачу определить орбиту хотя бы ближайшего к нам небесного тела, спутника нашей планеты — Луны: слишком велик был объем необходимых вычислений. Они не могли не досадовать по этому поводу, поскольку в те времена необходимость в определении траектории Луны вызывалась насущными потребностями практики. Тот, кто сумел бы вычислить траекторию Луны, одновременно решил бы важную проблему: точные таблицы движения Луны позволили бы капитанам определять местонахождение судна в открытом море. Такие таблицы положения небесных тел появились лишь недавно и были составлены при помощи новых вычислительных машин. Теория движения небесных тел проста, уравнения, описывающие их движение, не так уж сложны, но численное решение этих уравнений требует проведения миллионов арифметических операций. В прошлом веке над решением подобной задачи сотни людей должны были бы трудиться на протяжении нескольких лет. Ныне ту же задачу вычислительная машина успевает решить за несколько дней или недель.

Не следует думать, будто мы выбрали наиболее сложную задачу: небесная механика — еще сравнительно простая область. Движения воздушных масс в атмосфере, приливы и отливы, обтекание потоком корпуса судна или крыла самолета, волны, испускаемые атомом или радиоантенной, вопросы образования цен или движения капиталов в какой-нибудь отрасли экономики — все это гораздо более сложные математические задачи. Прежде чем приступить к решению, такие задачи значительно упрощают, но и облегченный вариант требует столь большого объема вычислительных работ, что для их выполнения приходится строить все более грандиозные и быстродействующие машины.

Мы хотим точно знать, какая погода будет завтра. В принципе предсказание погоды — задача вполне разрешимая. Располагая большими вычислительными машинами, ее можно решить не только «в принципе», но и численно. Однако нам придется запастись терпением: решение задачи потребует около года непрерывных вычислений!

Когда полвека назад голландцы решили перекрыть плотиной Зёйдер-Зе, им, естественно, понадобилось знать, какой высоты и формы должна быть дамба. Построив уменьшенную копию Зёйдер-Зе и модели различных плотин и воспроизведя в уменьшенном масштабе движения ветра и воды, исследователи занялись выяснением того, как сказываются на теле плотины приливы и отливы. Возможности вычислительных машин в то время были не столь широки, чтобы они могли оказать существенную помощь в подобных исследованиях. В наши дни при разработке более грандиозного по замыслу проекта «Дельта» по перекрыванию всех рукавов в устье Рейна без применения вычислительных машин уже не обойтись. Лишь численные методы помогут найти ответы на самые разнообразные вопросы — от создания схемы течений в огромной открытой чаше, называемой Северным морем, до выяснения особенностей поведения все более стремительных потоков в суживающемся по мере перекрытия русле того или иного рукава рейнской дельты.

Ныне во многих странах мира специальные бюро проводят расчеты, позволяющие составлять на ближайший год прогнозы развития экономики, производства, рынка труда, цен, заработной платы, потребления, капиталовложений и предсказывать последствия тех или иных мер, предпринимаемых правительственными органами. Такие расчеты можно было бы производить и раньше, но сколько времени понадобилось бы для получения ответа на любой из перечисленных выше вопросов? К 1980 г. мы могли бы совершенно точно предсказать, как следовало бы поступить в 1965 г. Лишь с появлением быстродействующих вычислительных машин решение столь сложных проблем приобрело практическое значение.

По отметке на экране радиолокационной станции можно определить расстояние до космического корабля, но не направление на него. Тем не менее, зная, как изменяется расстояние со временем, нетрудно вычислить орбиту космического корабля. Для того чтобы космонавтам дать команду на посадку, о включении тормозной двигательной установки и т. п., нужно точно знать, где находится космический корабль и какова его скорость в интересующий нас момент времени. Необходимые для этого вычисления наземная станция слежения должна производить за долю секунды. Столь же быстро следует определять момент, когда должен открыться затвор камеры межпланетной станции для фотографирования Луны или Марса.

Мы уже задали вопрос: а нужно ли нам торопиться? Поспешность — понятие относительное. Если на разработку плана «Дельта» потребовалось бы сто лет, то от такого проекта нам лучше было бы отказаться. Если составляется государственный план на 1967 г., то министр финансов не должен говорить: «Подождите до 1980 г., тогда я смогу вам сказать, какой налог с оборота целесообразно установить в 1967 г.» И космонавтам, ведущим счет времени на тысячные доли секунды, наземный центр космической связи не может ответить: «Подождите минуточку, пока мы тут все рассчитаем».

Автоматы

Во второй главе мы предложили изящный способ, позволяющий сколь угодно точно вычислить $\sqrt{2}$. Состоит он в следующем. Начав с довольно плохого приближения $(a_1 = 1)$, мы вычисляем одно за другим числа

$$b_1 = \frac{2}{a_1},$$

$$a_2 = \frac{1}{2}(a_1 + b_1),$$

$$b_2 = \frac{2}{a_2},$$

$$a_3 = \frac{1}{2}(a_2 + b_2)$$

и т. д.

и получаем все более узкие границы, между которыми заключено точное значение $\sqrt{2}$.

Почти так же выглядит и один весьма распространенный метод решения уравнений.

Записав уравнение в виде

$$f(x) = x,$$

где x — неизвестное, мы найдем грубое приближение значения корня x_1, подставим его в левую часть уравнения и получим x_2:

$$f(x_1) = x_2.$$

Аналогично поступим с новым приближенным значением корня x_2 и найдем его следующее приближенное значение x_3:

$$f(x_2) = x_3.$$

Вообще

$$f(x_n) = x_{n+1},$$

и если левая часть уравнения удовлетворяет определенным условиям (в случае необходимости ее можно преобразовать так, чтобы она удовлетворяла нужным условиям), то последовательные значения x_n будут все меньше отличаться друг от друга и служить все более точным решением уравнения

$$f(x) = x.$$

Задача математика состоит в том, чтобы разработать метод, позволяющий приближенно, но сколь угодно точно вычислять $\sqrt{2}$ и корни гораздо более сложных уравнений. Проведение самих расчетов математик может затем поручить вычислителю. Вычислитель должен знать таблицу умножения, а если в его распоряжении имеется вычислительная машина, то математик дает вычислителю следующую инструкцию (называемую также программой): «Ввести число a в регистр II (делитель), число 2 — в регистр I (делимое) и нажать кнопку D (деление). Затем стереть число в регистре I, перенести в этот регистр число a/b из регистра III и нажать кнопку C (сложение). Стереть числа в регистрах I и II, перенести число $a + 2/a$ из регистра III в регистр I и ввести в регистр II число 2. Нажать на кнопку D. Число, находящееся в регистре III и равное $^1/_2(a + 2/a)$, обозначить a, после чего повторить все операции с новым значением a. Вычисления прекратить, когда очередное значение a будет отличаться от предыдущего не более чем в последнем знаке (то есть по достижении требуемой точности)».

Для вычислителя не составляет особого труда точно выполнить все наставления математика, поскольку в действительности вычислитель умеет не только крутить ручку арифмометра и нажимать кнопки, но и производить несравненно более сложные операции. Инструкция вполне понятна ему, и в любой момент времени он знает, что следует делать для того, чтобы точно и неукоснительно выполнить все требования инструкции. Если мы хотим доверить выполнение инструкции машине, то машина должна уметь гораздо больше, чем только считать. Она должна «понимать» программу, последовательно выполнять ее и самостоятельно останавливаться, как только будет достигнут требуемый результат. С машинами, способными выполнять столь сложные действия, мы неоднократно встречались в повседневной жизни. Таков, например, автомат, продающий почтовые марки; утюг с терморегулятором, способный самостоятельно включаться и выключаться, стоит только температуре выйти за установленные пределы; проигрыватель, который автоматически выключается, когда запись на пластинке доходит до конца, и даже может перевернуть или сменить пластинку. Одним из первых автоматов такого рода стал ткацкий станок, изобретенный французом Ж. Жаккаром (1752—1834), для изготовления тканей со сложной схемой переплетения нитей основы и утка. Мы не встретим ни одного ткача, который бы вручную по заранее заданной схеме раздвигал нити основы и пропускал между ними челнок с нитями утка; работой автоматических ткацких станков в конечном счете управляют перфокарты. Встретив отверстие, пробитое в определенном месте перфокарты, рычажок приходит в движение и располагает нити основы и утка в требуемой комбинации.

Управляющее устройство и память

Вычислительная машина, производящая расчеты автоматически, помимо *арифметического устройства*, должна обладать *управляющим устройством*, задающим и контролирующим порядок выполнения операций, и *памятью*, хранящей исходные числовые данные, программу и промежуточные результаты до тех пор, пока не возникнет необходимость в их повторном использовании. Мысль о создании автоматической вычислительной машины впервые высказал англичанин Ч. Бэббедж (1792—1871), но после напряженной работы ему удалось создать лишь экспонат, украшающий ныне экспозицию музея в Кенсингтоне. Планам Бэббеджа не суждено было сбыться: они опередили свое время без малого на целое столетие. Общий уровень развития техники того времени (и, в частности, электротехники, делавшей лишь первые робкие шаги) не позволил осуществить Бэббеджу его дерзкий замысел. Гораздо больший успех выпал на долю американца Г. Холлерита (1860—1929), отца работающих на перфокартах вычислительных машин, без которых в настоящее время не обходится ни одно крупное предприятие. (Впрочем, нельзя не отметить, что планы

Г. Холлерита были несравненно определеннее и реалистичнее.) Но подлинный перелом в конструировании и производстве вычислительных машин произошел лишь после возникновения электронной промышленности. Одна за другой стали появляться конструкции вычислительных машин, и каждая следующая превосходила предыдущую — превосходила не только по размерам, но и по мощности, быстродействию и объему памяти. Современные вычислительные машины-великаны, выполненные на полупроводниках и печатных схемах, по своим размерам значительно меньше вычислительных машин-карликов двадцатилетней давности.

Для того чтобы вычислительная машина заработала, ее необходимо снабдить программой. Предположим, что требуется вычислить произведение 24 789 $\times$ $\times$ 345. Если мы работаем на неавтоматической вычислительной машине, то необходимо установить на регистре 24 789. Повернув ручку на 5 оборотов, мы должны перенести полученное число в регистр результатов и сдвинуть его на один разряд вправо. Затем необходимо установить на регистре 24 789 и, повернув ручку на 4 оборота, перенести полученное число в регистр результатов, где, таким образом, уже будет храниться произведение 24 789·45, после чего сдвинуть это произведение на один разряд вправо. Наконец, необходимо установить на регистре 24 789 и, повернув ручку на 3 оборота, перенести полученное число в регистр результатов. После чего там окажется записанным произведение 24 789·345.

Электромеханические вычислительные машины работают гораздо более «самостоятельно»: необходимо лишь установить на двух регистрах оба сомножителя (24 789 и 345) и затем нажать кнопку со знаком $\times$ (умножение). Все остальные операции (те же, что и в случае механической вычислительной машины) производит сама машина.

Совершенно иначе находит произведение электронная вычислительная машина. В ней нет ни регистров, ни десятичных счетных колес, ни рычагов, ни цифр, которые нужно устанавливать на колесах регистра или считывать. В машину лишь вводят перфоленту или колоду перфокарт, нажимают кнопку, и электрическая пишущая машинка выходного устройства напечатает конечный результат на длинной ленте.

На перфокартах «записано» все, о чем мы хотели бы рассказать машине: числовые данные, которые необходимы для вычислений, и подробная инструкция относительно того, какие операции необходимо произвести над числами для получения результата. Среди введенных в машину данных могут быть даже целые таблицы логарифмов, которые понадобятся машине, когда ей придется вычислять логарифмы каких-то чисел. Среди введенных инструкций может содержаться подробное описание метода извлечения квадратного корня или решения системы n линейных уравнений с n неизвестными.

Более того, инструкции могут включать описания даже двух методов решения одной и той же задачи, а машина уже должна будет, применяясь к обстоятельствам, выбирать наиболее подходящий из них.

Все, что машина считывает с перфокарт, она хранит в своей памяти. Каждому из нас случалось пользоваться магнитофоном. Мы знаем, что на магнитофонной пленке можно записать речь или музыку, а затем хранить и воспроизводить запись до тех пор, пока она не сотрется. Нечто похожее мы обнаружили бы, заглянув в электронную вычислительную машину: быстро вращающийся магнитный барабан, с которого машина за доли секунды может считать все, что на нем записано. На магнитном барабане записывают программу. Там же машина хранит и промежуточные результаты, которые могут понадобиться ей в дальнейшем. Кроме того, для результатов, которые могут понадобиться особенно срочно (их поступление машина может ожидать не сотые, а самое большее стотысячные доли секунды), в машине предусмотрено другое хранилище — так называемая «быстрая память». Запись в быстрой памяти и считывание хранящихся в ней данных производит электронный луч, а само хранилище срочной информации представляет собой катодную трубку, напоминающую кинескоп телевизора.

Память разделена на ячейки, а каждая ячейка снабжена номером — адресом. Предположим, что в программе стоит команда: «Взять число из ячейки

3456, умножить его на 2, а результат записать в ячейку 218» (или, быть может, любую другую свободную ячейку, которую машина найдет сама, а ее адрес запишет в другую заранее указанную ячейку).

Конечный результат вычислений машина записывает в памяти на магнитном барабане, а затем либо отпечатывает его на электрической пишущей машинке, либо вычерчивает в виде графика. Делает она это гораздо медленнее, чем производит вычисления. Выходное устройство может часами печатать результаты, поступающие с магнитного барабана, хотя сами вычисления давно закончились и машина занялась решением других задач. Над длинными бумажными лентами, испещренными колонками цифр, с интересом склоняются физик и астроном, биолог и экономист, ученый и инженер. Они хотят знать, что «сказала» машина, согласуются ли расчеты с результатами наблюдений и достаточно ли удовлетворительно объясняет созданная ими математическая теория процессы, происходящие в природе и обществе.

Язык отверстий

О чем могут сообщить машине перфокарты или перфоленты? Автоматическому ткацкому станку Жаккара перфокарты говорят, в какой последовательности должны подниматься или опускаться рычажки для того, чтобы нити основы и утка переплелись, образуя нужный узор. Так называемой пианоле — автоматическому пианино — перфокарты «подсказывают», в какой последовательности и с какой быстротой молоточки должны ударять по струнам.

Перфолента представляет собой узкую полоску бумаги с пробитыми в ней отверстями. В каждой строке умещается 5 отверстий. В одной строке бывают пробиты все 5 отверстий, в другой — ни одного. В строке может быть и 1, и 2, и 3, и 4 пробитых отверстия, причем строки с одинаковым числом отверстий могут отличаться их местоположением. Сколько различных сообщений способна вместить одна строка перфоленты? Условимся обозначать отверстие единицей, а его отсутствие — нулем. Тогда в одной строке нам могут встретиться следующие комбинации нулей и единиц:

0 0 0 0 0

0 0 0 0 1

0 0 0 1 0

0 0 0 1 1

0 0 1 0 0

0 0 1 0 1

и так далее, пока наконец не дойдем до комбинаций

1 1 1 0 0

1 1 1 0 1

1 1 1 1 0

1 1 1 1 1

Что напоминают нам такие комбинации? Ну конечно, не что иное, как записанные в двоичной системе числа от 0 до 31. Итак, одна строка перфоленты вмещает $31 = 2^5 - 1$ вариантов сообщений.

Десятичное число 99 999 999 в двоичной системе выглядит следующим образом:

000101111101011110000011111111.

Таким образом, самое большое восьмизначное число в десятичной системе можно записать в виде двадцатисемизначного числа в двоичной системе. На языке отверстий такое число можно записать, затратив $31 = 2^5 - 1$ вариантов сообщений.

Перевод на язык отверстий допускают не только числа. Перенумеровав буквы алфавита и заменив каждую из них соответствующим числом, можно записать буквы в нескольких строках перфоленты в виде определенной комбинации «есть отверстие» и «нет отверстия». К 26 буквам латинского алфавита, к основным символам, необходимо присоединить еще пробел между словами, точку, запятую, точку с запятой и так далее.

«Есть отверстие — нет отверстия» — лишь один из многих способов конкретного представления абстрактных нулей и единиц. Перфолента движется между зажимами пяти контактов, играя роль изолятора. Если в ней встречается отверстие, контакт замыкается и в цепи возникает ток. Таким образом, контакт осуществляет перевод с языка «есть отверстие — нет отверстия» на язык «есть ток — нет тока». Текущий по цепи ток проходит по обмотке электромагнита, расположенного вблизи магнитного барабана, поверхность которого усеяна кристалликами феррита. Электромагнит воздействует на кристаллики вращающегося барабана. Кристаллики, на которые электромагнит «смотрит» в тот момент, когда по его обмотке течет ток, намагничиваются и остаются в таком состоянии, то есть маленькими магнитиками, до тех пор, пока их не размагнитят. Таким образом, электромагнит занимается переводом с языка «есть ток — нет тока» на язык «намагниченный — ненамагниченный». При считывании информации, записанной на барабане, намагниченный участок барабана индуцирует электрический ток, который поступает к другому узлу машины, например к переключателю, где происходит перевод с языка «намагниченный — ненамагниченный» на язык «выключено — включено». Лампа накаливания перевела бы тот же «текст» на язык «темно — светло». На языке кинескопа то же самое означало бы «есть электрический заряд — нет электрического заряда». Реле восприняло бы сообщение, переданное с магнитного барабана, как «контакт замкнут — контакт разомкнут». Полупроводниковый вентиль перевел бы то, о чем ему поведал ток, по-своему: «есть проводимость — нет проводимости». Все узлы машины вплоть до конечного устройства, переводящего результат из двоичной системы в десятичную и выдающего его в десятичном виде на печатное устройство, говорят на языке «0 — 1» и понимают этот язык.

Первоначально перфоленты нашли использование не в вычислительных машинах, а в телеграфных аппаратах. Текст, записанный на перфоленте на языке «есть отверстие — нет отверстия», передавался по проводам или в эфир. Если в ленте отверстия не было, то не было и электрического сигнала. Отверстие в перфоленте соответствовало электрическому сигналу. Перфолента позволяет передавать по проводам и без проводов не только текст, но и изображения, если их предварительно перевести на язык нулей и единиц.

Как это делается? На фотографию размером 8×16 (размеры даны в сантиметрах) накладывается растр, разделенный на 64 части по ширине и на 128 частей по высоте. Фотография оказывается разделенной на $64 \cdot 128 = 2^{13}$ клеток, которые можно перенумеровать парами чисел: одно число означает номер клетки по ширине, другое — номер клетки по длине фотографии. Номерами клеток по ширине служат числа от 0 до 63, номерами по длине — числа от 0 до 127. Разумеется, оба номера следует записать в двоичной системе: номера по ширине принимают значения от 000000 до 111111, номера по длине — от 0000000 до 1111111. Пестрая последовательность 13 нулей и единиц позволяет перенумеровать клетки, на которые разбита фотография. Затем фотоэлемент «просматривает» (сканирует) одну за другой все клетки и определяет, какие из клеток темные и какие светлые. Именно эту информацию можно передать с помощью перфоленты: информация, поступающая на вход канала связи и снимаемая с выхода, по существу остается неизменной. Иначе обстоит дело с информацией, вводимой в вычислительную машину и извлекаемой из нее: вычислительная машина не просто доставляет информацию от входа к выходу, но и перерабатывает сообщенные ей данные.

Переработка информации

Переработка исходных данных происходит постепенно, мелкими шагами, каждый из которых длится одну стотысячную или даже одну миллионную секунды.

Сложение и умножение двух тридцатизначных чисел (в двоичной системе) представляет собой довольно сложную операцию. Простейшие случаи сложения и умножения двух чисел сведены в таблицу двоичного сложения и умножения, о которой мы уже говорили:

$$0+0=00, \quad 0+1=01, \quad 1+0=01, \quad 1+1=10,$$
$$0 \cdot 0=0, \quad 0 \cdot 1=0, \quad 1 \cdot 0=0, \quad 1 \cdot 1=1.$$

Вычисляя произведение двух чисел a и b, каждое из которых принимает лишь значения 0 или 1, мы должны найти число c, также принимающее значения 0 или 1. Вычисляя сумму $a+b$ тех же чисел, мы можем представить ее в виде «двузначного» двоичного числа cd, где c — цифра, стоящая слева, а d — цифра, стоящая справа. Все возможные комбинации значений a, b, c, d перечислены в следующей таблице:

a	b	c	d
0	0	0	0
0	1	0	1
1	0	0	1
1	1	1	0

Вместо единиц и нулей условимся снова говорить «ток есть» и «тока нет». Тогда числам a, b, c, d будут соответствовать какие-то проводники. Если по проводникам a и b течет ток, то по проводнику c также

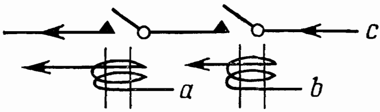


Рис. 26.

течет ток. Но если хотя бы по одному из проводников a или b ток не течет, то не течет ток и по проводнику c. Невольно напрашивается мысль о последовательном соединении: в цепи c имеются два контакта, управляемые, например, реле a и b (рис. 26). Когда по обмоткам реле a и b течет ток, электромагниты втягивают сердечники, замыкают контакты и по цепи c течет ток. Но если хотя бы в одной из обмоток a или b тока нет, то по крайней мере один из контактов разомкнут и по цепи c ток не идет.

Сложнее обстоит дело с проводником d. Если проводники a и b находятся в одном и том же состоянии (то есть либо по a и по b идет ток, либо по a не идет ток и по b также не идет ток), то и по d ток не идет. Если же a и b находятся в различных состояниях (то есть по одному из проводников идет ток, а другой, как говорят специалисты, «обесточен»), то по d ток идет. Кому не придет в голову мысль о лампе, включаемой из двух различных мест? Если оба выключателя находятся в одинаковом положении (либо «вкл.», либо «выкл.»), то лампа потушена. Если выключатели находятся в различных положениях, то лампа горит.

На рис. 27 показана релейная схема, аналогичная схеме включения лампы из двух различных мест: оба выключателя находятся в одинаковом состоянии, поэтому цепь d разомкнута и ток по ней не идет. Если же по одной из обмоток a или b идет ток, то срабатывает лишь один электромагнит и положение «своего» контакта изменяет на противоположное; цепь d замыкается. Если же ток течет и по обмотке a, и по

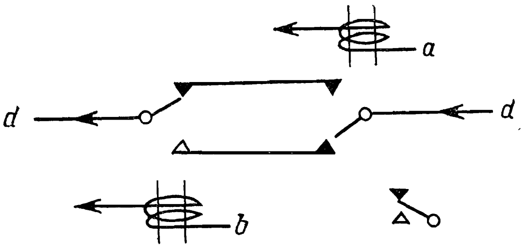


Рис. 27.

обмотке b, то срабатывают оба электромагнита, оба контакта изменяют свое положение, и цепь d снова оказывается разомкнутой.

Из двух величин a и b, принимающих лишь значения 0 или 1, новую величину x, также принимающую лишь значения 0 или 1, можно записать четырьмя способами. Таблица

a	b	x
0	0	x_1
0	1	x_2
1	0	x_3
1	1	x_4

показывает, что величина x принимает значения x_1, x_2, x_3, x_4, когда ab означает пары значений 00, 01, 10, 11. Существуют четыре арифметических действия. Следовательно, двум известным величинам a, b новую величину x можно сопоставить 16 различными способами, и каждый из них допускает представление в виде релейно-контактной схемы. Впрочем, никто не заставляет нас начинать сразу с двух известных величин a и b. Мы можем взять лишь одну известную величину a, принимающую лишь два значения 0 или 1, и сопоставить ей новую величину x, также принимающую лишь два значения 0 или 1:

a	x
0	x_1
1	x_2

Сделать это можно четырьмя способами: значения x либо совпадают со значениями a, либо противоположны им (то есть x принимает значение 1 при $a = 0$ и $x = 0$ при $a = 1$).

Истинно и ложно

Какие выгоды сулит нам введение новой величины, станет ясно, если мы условимся переводить 0 как «ложно», 1 как «истинно» и понимать под a, b, ..., x высказывания, которые могут быть либо истинными, либо ложными? При переработке информации из двух таких высказываний возникает одно новое составное высказывание. Рассмотрим, например, следующее высказывание:

a: «1 декабря — не воскресенье».
b: «2 декабря — не воскресенье».
x: «1 декабря — не воскресенье или 2 декабря — не воскресенье».
y: «1 декабря — не воскресенье и 2 декабря — не воскресенье».
z: «Если 1 декабря — не воскресенье, то 2 декабря — не воскресенье».

Высказывание x всегда истинно, независимо от того, истинны или ложны высказывания a и b, поскольку из двух последовательных дат по крайней мере одна приходится не на воскресенье. (Условимся всегда считать составное высказывание с союзом «или» истинным, если истинны либо одно из высказываний a или b, либо оба высказывания.) Наоборот, высказывание y в одни годы истинно, в другие ложно. Так же обстоит дело и с высказыванием z. Оно ложно, если 1 декабря не приходится на воскресенье, а 2 декабря — воскресенье, то есть в те годы, когда 1 декабря падает на субботу.

Еще один пример.

a: «1 декабря — воскресенье».
b: «Рождество падает на 1 декабря».
c: «Рождество падает на 25 декабря».
d: «Новый год приходится на 1 января».
e: «Пасха приходится на 2 января».

В этом случае составные высказывания

c и d,
a или c,
d или e,
если b, то c,
если b, то e

всегда истинны, составные высказывания

d и c,
b и e,
b или e,
если c, то e

всегда ложны, а составные высказывания

a или b,
a и c,
если a, то b

могут иногда быть истинными, иногда ложными (последнее высказывание ложно в те годы, когда 1 декабря — воскресенье, поскольку в эти годы, если бы последнее высказывание было истинным, рождество приходилось бы на 1 декабря).

Предположим теперь, что u и v — два высказывания, относительно которых известно лишь, истинны они или ложны. Этого достаточно, чтобы мы могли определить, истинны или ложны составные высказывания «u и v», «u или v», «если u, то v». Действительно, высказывание «u и v» истинно лишь в том случае, если оба высказывания u и v истинны одновременно. Высказывание «u или v» ложно лишь в том случае, если оба высказывания u и v одновременно ложны. Наконец, высказывание «если u, то v» ложно лишь в том случае, если u истинно, а v, наоборот, ложно.

Вернемся вновь к нулям и единицам: условимся ставить 1 всюду, где стоит «истинно», и 0 всюду, где стоит «ложно». Значениям u и v связки «и», «или», «если ..., то ...» сопоставляют новые значения, которые приведены в следующей таблице:

u	v	u и v	u или v	если u, то v
0	0	0	0	1
0	1	0	1	1
1	0	0	1	0
1	1	1	1	1

Мы видим, что союз «и» действует так же, как суффикс «-жды» при умножении (дважды два — четыре), поэтому «и» можно представить при помощи той же

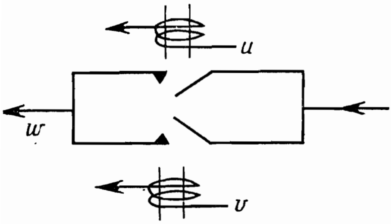


Рис. 28.

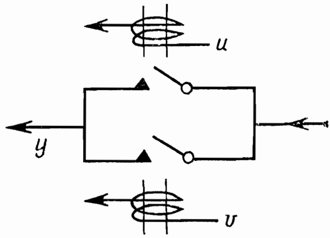


Рис. 29.

релейно-контактной схемы. Взглянув на столбец, описывающий действие «или», нетрудно заметить, что «или» реализуется при параллельном соединении (рис. 28): ток в цепи «u или v» не течет лишь в том случае, когда нет тока и в обмотке u, и в обмотке v.

Релейно-контактная схема, реализующая «если u, то v» изображена на рис. 29: ток в цепи «если u, то v» не течет лишь в том случае, если в обмотке u есть ток, а в обмотке v тока нет.

Релейно-контактную схему, изображенную на рис. 27, также можно перевести на язык «истинно — ложно»: d означает, что высказывания a и b противоречивы.

Совсем не обязательно преобразовывать два высказывания в одно составное: взяв лишь одно высказывание, мы также можем получить из него новое высказывание — при помощи операции отрицания исходного высказывания. Высказывание «не *и*» образуется из высказывания *и* следующим образом:

и	не *и*
0	1
1	0

Релейно-контактная схема, преобразующая *и* в «не *и*», показана на рис. 30: если по обмотке *и* не течет ток, то по цепи «не *и*» идет ток (контакт замкнут); если же по обмотке *и* идет ток, то в цепи «не *и*» тока нет (контакт разомкнут).

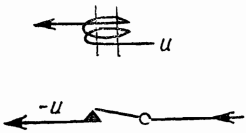


Рис. 30.

Слова «и», «или», «если ..., то ...», «не» и так далее называются *логическими связками*, поскольку они не являются самостоятельными высказываниями, а лишь связывают истинность и ложность. Логика уделяет много внимания изучению мышления, и позднее мы еще поговорим о думающих машинах. Но сначала нам необходимо познакомиться с автоматами, умеющими играть в различные игры.

Играющие машины

Теперь уже, должно быть, все читатели догадались, почему мы так подробно разбирали игру в ним. Правила игры в ним и секрет, позволяющий здесь быть чемпионом, столь просты, что их можно объяснить даже весьма примитивной вычислительной машине. Какими «способностями» должна обладать машина для того, чтобы наши объяснения оказались доступными для нее? Она должна уметь переводить числа, заданные ей в десятичной системе, в двоичную систему, подсчитывать число единиц в столбцах и определять, четно оно или нечетно, и, если возможно, уменьшать одно из чисел так, чтобы число единиц в каждом столбце становилось четным. Каждый, кто хоть немного разбирается в технике и знаком с релейно-контактными схемами, без труда набросает схему такой машины, не требующей особых затрат при постройке.

Одна машина, играющая в ним, демонстрировалась на Всемирной выставке в Чикаго вскоре после окончания второй мировой войны (позднее эту машину показывали и в Европе). Весила она около тонны, но это была лишь видимость: в действительности машину можно было переносить в ящике от большого радиоприемника. Машина вела себя по-джентльменски: она всегда предлагала своим партнерам исходные позиции, которые поднимали их шансы на выигрыш. И все же в подавляющем большинстве партий победу одерживала машина. Ее партнеры либо не владели секретом беспроигрышной игры в ним, либо, если и были осведомлены о нем, не могли состязаться с машиной в быстроте счета и безошибочности выкладок. Импрессарио непревзойденного игрока в ним вовремя заметил, что сверхчеловеческая быстрота, с которой машина производит необходимые для очередного хода расчеты, угнетающе действует на зрителей, и, чтобы предотвратить провал задуманного им предприятия, искусственно «притормозил» машину. На каждый ход своего партнера машина отвечала не сразу, а несколько помедлив, как бы после обдумывания создавшейся позиции.

Говоря об игральных автоматах, прежде всего имеют в виду автоматы для различных азартных игр и лишь потом вспоминают о машинах, умеющих играть в них на гроссмейстерском уровне. Такие азартные игры, как орел или решка, кости, рулетка, автоматизированы лишь недавно. При очень длинном ряде бросаний нефальшивая монета будет выпадать вверх гербом столь же часто, как и вверх решеткой. В этом можно не сомневаться. Зато все остальные детали поведения монеты нельзя предсказать с полной уверенностью, и именно в такой неопределенности и заключена «соль» всех азартных игр. Даже пронаблюдав за длинной серией бросаний монеты, нельзя предсказать, выпадет ли она при следующем бросании вверх гербом или вверх решеткой. Аналогично обстоит дело и с игральными костями или с рулеткой. Стоит лишь появиться какой-нибудь закономерности между начальным и конечным положением колеса рулетки в игральном зале Монте-Карло, как такую рулетку тотчас же отправят в ремонт.

Во многих государствах азартные игры, любители которых подчас проигрывают баснословные суммы, запрещены законом. Владельцы игральных автоматов утверждают, будто исход подобных игр во многом зависит от искусства игрока. Нельзя сказать, что подобные заявления полностью не соответствуют действительности. Чтобы привести игральный автомат в действие, достаточно мелкой монеты. Автоматы, как правило, очень изношены, и, экспериментируя, нетрудно подметить какую-нибудь закономерность в их работе и даже обнаружить возможность влиять на «ходы» автомата. Например, при игре на автомате, подвергнутом исследованию, играющий с вероятностью 60% мог вернуть свои деньги, а при умелой игре его шансы на возвращение затраченных на игру денег возрастали до 70%. Иначе говоря, на каждую опущенную в автомат марку играющий, действуя наугад, в среднем терял 40 пфеннигов. Проявив известную сноровку, он мог снизить потери до 30 пфеннигов на марку. В действительности самым искусным следует считать того игрока, который вообще не играет, поскольку его средний проигрыш заведомо равен нулю. Впрочем, уже неоднократно высказывались предложения вместо того, чтобы вести безнадежную борьбу с игральными автоматами, обязать их владельцев вывешивать у каждого автомата объявление, где бы в явном виде говорилось, каковы в среднем потери при игре наугад и при искусной игре. Это было бы наиболее действенным средством борьбы с той напастью, которую представляют игральные автоматы.

Игра в чет и нечет

А сейчас речь пойдет об автоматах совсем иного рода. Игра известна под названием чет и нечет. Играют в нее вдвоем: игрок *A* и игрок *B*. Оба одновременно показывают друг другу монетку. Если на лицевой стороне обеих монет оказывается герб, то выигрывает *A*. Если на лицевой стороне одной монеты оказывается герб, а на лицевой стороне другой монеты — решетка, то выигрывает *В*.

Сами по себе шансы игроков *А* и *В* на выигрыш равны. Но при игре в чет и нечет ловкость рук и сообразительность играют далеко не последнюю роль. Например, если *В* настолько глуп, что попеременно показывает герб и решетку, то игроку *А* для того, чтобы выиграть, достаточно лишь повторять ходы своего партнера. Если такую глупость совершает игрок *А*, то игрок *В* обеспечивает себе победу, повторяя все ходы противника «наоборот», то есть показывая другую сторону монеты, чем *А*. Разумеется, столь глупой тактики не придерживается никто. Но и небольшой оплошности достаточно, чтобы проиграть более хитрому противнику. Тот из игроков, кому удается раскрыть замысел своего партнера, получает преимущество. Отсюда следует, что лучше всего вообще не придерживаться определенного плана, а показывать герб и решетку наугад, случайным образом: ведь если у вас нет плана, то и раскрыть его невозможно; если играть бессистемно, не придерживаясь какой-либо схемы, то партнер бессилен воспользоваться обнаруженной им закономерностью в своих интересах. Однако достичь полной беспорядочности в ходах не так-то легко. Если игра продолжается достаточно долго, то игрок, сам того не замечая, начинает отдавать предпочтение тем или иным ходам — в распределении гербов и решеток обнаруживаются некоторые закономерности. Однажды нескольких людей попросили выписать наугад ряд чисел. И что же? В полученных последовательностях «случайных» чисел выявились вполне закономерные повторения отдельных групп чисел, связанных с днями рождения, номерами домов, телефонов и тому подобное.

Свойственной людям склонностью к закономерности некогда воспользовался К. Шеннон. Он построил машину, игравшую в чет и нечет с человеком и выигрывавшую в среднем 60% партий, хотя игра велась честно: шансы на выигрыш машины и игрока были равны. Вместо того чтобы показывать машине герб или решетку, человек вводил в нее 1 или 0. Как только человек нажимал соответствующую кнопку, машина сообщала свой ответный ход: показывала в окошке

0 или 1. Если оба партнера предъявляли друг другу одно и то же число, то выигрывала машина, в противном случае победителем объявлялся человек. Человек делал несколько ходов, а машина записывала все происходящее в своей памяти. Некоторое время спустя после начала игры машина анализировала накопленный опыт и поведение партнера, а затем использовала обнаруженные ею закономерности против него. Машине с ее прочной памятью и способностью молниеносно производить вычисления удавалось раскрывать тайные замыслы своих партнеров, которые те, сами того не сознавая, пытались претворить в жизнь. Обезопасить себя при игре с машиной можно было, лишь исключив «тайные» замыслы: например, партнер мог тайком бросать монету и в зависимости от исхода бросания нажимать либо кнопку «0», либо кнопку «1». Кстати сказать, именно так поступала машина Шеннона в начале игры, когда она еще недостаточно знала своего партнера и на всякий случай пыталась скрыть от него свои «замыслы».

Метод Монте-Карло

Итак, мы выяснили, что машина Шеннона должна уметь не только считать, но и играть в орла или решку и работать не только в закономерном, или, как еще говорят, детерминированном, но и в случайном режиме. Машина Шеннона должна владеть искусством производить случайные последовательности нулей и единиц. То же можно сказать не только о машинах, построенных для того, чтобы играть в различные игры. Существует немало практических проблем, настолько запутанных и сложных, что детерминированный подход к их решению сопряжен с большими трудностями или заведомо обречен на неудачу. При решении подобных задач вычислительной машине не остается ничего другого, как играть в случайную игру. Такой подход к решению задач называется *методом Монте-Карло.*

Рассмотрим один пример. Электрическая сеть рассчитана на определенную среднюю нагрузку, но ее нагрузка значительно возрастает в часы пик. Периоды повышенного расхода электроэнергии, как правило, хорошо известны: таковы, например, зимние вечера, когда промышленные предприятия еще продолжают работу, но уже включается освещение улиц, квартир, световая реклама. Вместе с тем наблюдаются и неожиданные пики расхода электроэнергии, вызванные стечением случайных обстоятельств, например внезапным наступлением холодов, заставляющих жителей города включать электрические обогреватели, и показом в тот же вечер сенсационной телепередачи, которую смотрит множество народа. Электростанция оказывается перегруженной, и напряжение в сети падает. Вычислить частоту подобных нежелательных перегрузок — задача математика. Его подход к решению мы рассмотрим на несколько более простом примере. Предположим, что нас интересует работа электростанции, обслуживающей обширную сеть электрофицированных железных дорог. Нежелательная перегрузка может возникнуть, если вследствие недосмотра в графике движения или случайного опоздания в один и тот же момент времени будут трогаться с места очень много поездов. Такого быть не должно, поэтому на большой электронной вычислительной машине заранее «проигрывают» график движения по всем участкам железнодорожной сети, причем в программу вводят всевозможные опоздания, распределенные по времени случайным образом, и рассчитывают возникающую при различных ситуациях перегрузку электростанции. Так численные методы позволяют изучать потребности в электроэнергии при пиках нагрузки и позволяют уже со знанием дела решать вопрос о выборе мощности проектируемой электростанции.

Проигрывание на машине «игрушечной» ситуации называется *моделированием*. Например, можно составить схему народного хозяйства, где одна отрасль производит, другая потребляет, одна продает, другая покупает, где отрасли взаимосвязаны и взаимозависимы, и учесть в программе различные случайные факторы. Машина, которой придется моделировать эту схему, должна уметь играть в случайные игры. Для этого ей не обязательно учиться играть в орла или решку. Вычислительные машины умеют производить хитроумнейшие случайные последовательности с заранее заданным распределением нулей и единиц.

Детерминированные вычислительные машины

Различают детерминированные вычислительные машины, перерабатывающие вводимый в них материал так, что каждый следующий шаг в превращении исходных данных в окончательный результат целиком определяется предыдущим шагом, и вероятностные вычислительные машины, в которых переработка вводимого материала зависит от случайного элемента. В детерминированных машинах конечный результат полностью определяется тем, что́ поступило на вход машины. Если ввести одни и те же исходные данные в две детерминированные машины-близнецы, то обе машины выдадут одинаковые цифры. Наоборот, если машина использует какие-то случайные элементы, то при одних и тех же исходных данных она будет выдавать то один, то другой результат, подобно тому как при вытягивании билетиков в лотерее невозможно два раза получить одинаковую последовательность номеров.

Следует заметить, что строго детерминированная вычислительная машина — не более чем абстрактное понятие. Ошибаться свойственно не только человеку. Машины также могут это делать. Современная вычислительная машина, выполняющая в секунду миллион операций, время от времени может «недоглядеть» и пропустить одну из них, контакт может не замкнуться, электронная лампа — не сработать, полупроводник — перегреться. Такие сбои сказываются, хотя и незначительно, на конечном результате и изменяют его, хотя исходные данные остаются неизменными. На практике не существует вычислительных машин, которые можно было бы назвать детерминированными в смысле данного выше определения, но представлять себе такую машину как абстракцию вполне допустимо.

Нельзя не отметить еще одно весьма примечательное обстоятельство: цифровые последовательности, порожденные вполне детерминированным образом, внешне могут выглядеть случайными последовательностями. Например, предположим, что машине задано восьмизначное число a и она должна умножить его на 23, разделить полученное произведение на $10^8 + 1$ и записать остаток в виде нового числа a. Если числа a

выписать одно за другим, то они образуют последовательность, в которой нельзя обнаружить никакой закономерности. Располагая запрограммированной таким образом машиной, можно открыть игорный дом и утверждать, что игра, предлагаемая вниманию посетителей, лишена элемента случайности. Выдаваемая машиной числовая последовательность образуется по вполне определенному закону, и тот, кто пожелает сорвать банк, должен лишь в течение некоторого времени пронаблюдать за чередованием цифр, чтобы заметить скрытую закономерность. И все же подобное утверждение следует расценить как мошенничество. Обнаружить скрытую закономерность чередования цифр можно лишь в том случае, если воспользоваться гораздо большей вычислительной машиной, чем та, которая производит последовательность. Обнаружить закон, по которому машина печатает очередной член последовательности, «невооруженному» человеку не под силу. Для него числовая последовательность столь же случайна, как исход запуска рулеточного колеса в Монте-Карло. Ему кажется, будто все цифры имеют одинаковые шансы на появление и все зависит лишь от случая.

Что делают детерминированные вычислительные машины?

Или (что гораздо важнее): чего не делают детерминированные вычислительные машины? На этот вопрос мы дадим строгий математический ответ, но предварительно нам необходимо уточнить, что́ следует понимать под детерминированной вычислительной машиной. Определение детерминированной вычислительной машины достаточно общего и вместе с тем достаточно простого типа впервые сформулировал Алан Тьюринг (1912—1954). В честь его ныне такие машины принято называть машинами Тьюринга.

Машина Тьюринга должна уметь читать и писать. Читает она бумажную ленту, разделенную на клетки, и пишет на этой же ленте. С одной стороны лента ограничена, с другой — простирается до бесконечности. В каждой клетке может стоять либо 0, либо 1. Машина знает эти цифры: умеет их читать и писать. Своим «глазом» она обозревает клетку 1, расположенную прямо по лучу зрения, и в зависимости от того, какая цифра вписана в эту клетку, командует ленте либо «вперед», либо «назад», чтобы обозреть следующую или предыдущую клетку. Машина может вписать в обозреваемую клетку 0 или 1, а также, стерев стоящий в клетке нуль, заменить его единицей, или, наоборот, стереть стоящую в клетке единицу и заменить ее нулем. Что предпримет машина, зависит не только от содержимого обозреваемой ею клетки, но и от «внутреннего состояния», в котором она находится. Машина Тьюринга может находиться в конечном числе таких внутренних состояний: $S_1, \ldots, S_N$.

Несколько упрощенно можно представить себе, что нули вообще не пишутся. Нулями служат пустые клетки. Таким образом, вписать нуль в клетку означает либо ничего не делать (если обозреваемая машиной клетка пуста), либо стереть стоящую в клетке единицу.

Машина Тьюринга может производить следующие действия:

1) вписать штрих в обозреваемую клетку (если в клетке стоит штрих, то машина оставляет его без изменений), условное обозначение операции 1;

2) стереть штрих в обозреваемой клетке (если клетка пуста, то машина оставляет ее без изменений), условное обозначение операции 0;

3) сдвинуть ленту на одну клетку дальше и обозреть следующую клетку, условное обозначение операции $+$;

4) сдвинуть ленту на одну клетку назад и обозреть предыдущую клетку, условное обозначение операции $-$;

5) прежде чем обозревать некоторую клетку, перейти в новое внутреннее состояние S_i; условное обозначение операции S_i.

Бумажная лента не ограничена в положительном направлении — вперед. В отрицательном направлении лента ограничена некоторой начальной клеткой. Прокручивая ленту назад, машина не может «перешагнуть» через начальную клетку и, дойдя до нее, останавливается. Когда машина приступает к работе, на бумажной ленте в конечном числе клеток уже могут стоять штрихи. Глаз машины направлен на первую клетку, сама машина находится во внутреннем состоянии S_1. В дальнейшем машина работает автоматически. Если машина после конечного числа шагов останавливается, то результат ее деятельности отражает лента, в определенные клетки которой вписано конечное число штрихов. Такой результат трудно назвать интересным: конечное число знаков можно написать и от руки, не прибегая к услугам автомата. Гораздо интереснее случай, когда машина работает без остановки. В этом случае она может постепенно образовать бесконечную последовательность нулей и единиц, например записать число π в двоичной системе. Особо подчеркнем, что машина *может* породить бесконечную последовательность нулей и единиц. Мы отнюдь не утверждаем, что машина непременно *должна* произвести такую последовательность. Нетрудно представить себе случай, когда машина, вписав в одну или в несколько клеток по единице, при последующих ходах сотрет все, что она вписала, затем снова впишет и снова сотрет и так до бесконечности. В этом случае мы не можем утверждать, что на бумажной ленте возникнет бесконечная числовая последовательность. Такое утверждение обосновано лишь тогда, когда в некоторый момент времени мы можем со всей определенностью установить содержимое любой клетки и узнать, что в ней написано: нуль или единица.

Пример I. Машина находится во внутреннем состоянии S. Ее реакцию на «внешние раздражители» можно записать в следующем виде:

$$S0 \to +1S, \quad S1 \to +0S.$$

Это означает, что, увидев в обозреваемой клетке нуль, машина впишет в следующую клетку единицу, а увидев в обозреваемой клетке единицу, — впишет в следующую клетку нуль. Если до начала работы машины бумажная лента пуста, то машина, действуя автоматически, порождает бесконечную последовательность

$$010101 \ldots .$$

Если же в первой клетке стоит единица, то машина порождает бесконечную последовательность

$$101010\ldots .$$

Пример II. Рассмотрим машину, повторяющую заданную конечную числовую последовательность в любом указанном ей месте ленты (например, записывающую последовательность-копию сразу после окончания последовательности-оригинала). Предположим, что исходная последовательность, содержащая n знаков, записана в клетках с четными номерами $2, 4, \ldots, 2n$ и нам требуется воспроизвести ее копию в клетках с номерами $2n+2, \ldots, 4n$. Клетки с нечетными порядковыми номерами мы по мере надобности будем использовать для «черновых» выкладок. В обе клетки с нечетными номерами, непосредственно примыкающие к тем клеткам, в которых расквартированы члены заданной последовательности, то есть в клетки с номерами 1 и $2n+1$, впишем по единице, означающей начало и конец.

Машина может находиться в семи внутренних состояниях: $S_1, \ldots, S_7$. Ее реакция на «внешние раздражители» задана соотношениями

$$S_1 0 \to --S_1,$$
$$S_1 1 \to ++1-S_2,$$
$$S_2 0 \to +++S_3, \qquad S_2 1 \to +++S_4,$$
$$S_3 0 \to ++S_3, \qquad S_4 0 \to ++S_4,$$
$$S_3 1 \to ++S_5, \qquad S_4 1 \to ++S_6,$$
$$S_5 0 \to 1-0-S_7, \qquad S_6 0 \to 1-1-S_7,$$
$$S_5 1 \to ++S_5, \qquad S_6 1 \to ++S_6,$$
$$S_7 0 \to --S_1,$$
$$S_7 1 \to --S_7.$$

В начальный момент машина находится во внутреннем состоянии S_1 и обозревает штрих в первой клетке. На $S_1 1$ она отвечает ходом $++1-S_2$, то есть вписывает единицу в третью клетку, обозревает вторую клетку и переходит во внутреннее состояние S_2. Ознакомившись с содержимым второй клетки, машина «вглядывается» в пятую клетку и при этом переходит в состояние S_3, если во второй клетке стоял нуль, и в состоянии S_4, если во второй клетке стояла единица. Тем самым машина запоминает содержимое второй клетки. Нуль, стоящий в пятой клетке, означает, что машина еще не дошла до конца заданной последовательности. На S_30 и на S_40 машина реагирует, возвращаясь к седьмой клетке и сохраняя неизменным свое внутреннее состояние. То же самое повторяется и в седьмой, и в девятой и во всех остальных клетках с нечетными номерами до тех пор, пока машина не дойдет до клетки, в которую вписана единица, означающая, что машина дошла до конца заданной последовательности. На S_31 и S_41 в $(2n+1)$-й клетке машина отвечает продвижением на 2 клетки вперед, переходя на этот раз во внутреннее состояние S_5 (из S_3) или S_6 (из S_4). В $(2n+3)$-й клетке стоит нуль (клетка пуста). Обнаружив его, машина совершает в зависимости от того, в каком из двух внутренних состояний она находится, либо ход $1 - 0 - S_7$, либо ход $1 - 1 - S_7$. Иначе говоря, машина вписывает в $(2n+3)$-ю клетку единицу, которая понадобится в дальнейшем, отступает на одну клетку назад, переносит в $(2n+2)$-ю клетку содержимое второй клетки и, отступив назад еще на одну клетку, переходит в состояние S_7. В $(2n+1)$-й клетке машина обнаруживает единицу. На S_71 она отвечает, отступая назад на 2 клетки, после чего в поле зрения машины оказывается $(2n-1)$-я клетка, а сама машина по-прежнему остается во внутреннем состоянии S_7. В $(2n-1)$-й клетке она, вообще говоря, обнаруживает нуль, отступает назад еще на 2 клетки, но при этом переходит во внутреннее состояние S_1. Так машина «пятится» до тех пор, пока, наконец, не доходит до единицы. Это произойдет, лишь когда машина доберется до третьей клетки. Затем она повторит все операции, чтобы перенести в $(2n+3)$-ю клетку содержимое третьей клетки. Так будет продолжаться до тех пор, пока во все промежуточные клетки с нечетными номерами машина не впишет единицы. Нетрудно убедиться, что затем машина безостановочно пятится и, наконец, доходит до первой клетки, после чего останавливается: копия заданной конечной последовательности построена.

Нашу машину нетрудно было бы усовершенствовать таким образом, чтобы по окончании работы она стирала все единицы, вписанные ею в клетки с нечетными номерами.

По-видимому, двух приведенных примеров достаточно для того, чтобы читатель уверовал в способность машин Тьюринга выполнять четыре арифметических действия, извлекать корни и решать уравнения n-й степени с целыми коэффициентами, то есть выписывать решения таких уравнений в виде бесконечных непрерывных дробей. Коэффициенты уравнения можно ввести в машину, представив их в виде двоичных чисел на бумажной ленте; границу между коэффициентами можно указывать условной меткой. Положительность или отрицательность коэффициентов можно задавать, например, вписывая две условные метки в клетки с нечетными номерами, а под запись самих коэффициентов отвести лишь клетки с четными номерами.

Нетрудно также указать «конструкцию» машины Тьюринга, разлагающей любое вводимое в нее натуральное число на простые множители, машины, вписывающей двоичное разложение числа π или e, или машины, вычисляющей логарифм целого или алгебраического числа, то есть разлагающей логарифм в бесконечную непрерывную дробь. К этой же категории машин относится и машина, вычисляющая таблицы логарифмов (например, двадцатизначные логарифмы целых чисел).

Можно спросить: найдется ли какая-нибудь задача, которая была бы не по силам машинам Тьюринга? На этот вопрос мы ответим чуть позже, а пока зададим еще один вопрос: можно ли считать машины Тьюринга самым общим типом детерминированных машин, какой только существует? Казалось бы, что придумать машины «более общего типа» совсем нетрудно: таковы, например, машины, умеющие читать и писать не только нули и единицы, но и другие знаки, машины с несколькими бумажными лентами, несколькими «глазами» (считывающими устройствами) и так далее.

Такие машины могут оказаться более удобными в работе и быстрее производить необходимые действия, но все, что сумеют сделать они, сумеют сделать и машины Тьюринга. Любой алфавит с конечным числом знаков всегда можно заменить алфавитом, состоящим лишь из двух знаков: 0 и 1. Все, что записано на нескольких лентах и прочитано несколькими «глазами», можно записать на одной ленте и прочитать одним глазом. Точного определения того, что следует называть детерминированной вычислительной машиной в общем случае, не существует. Не имея точного определения, нельзя доказать, что машины Тьюринга «всемогущи», то есть умеют выполнять все операции, выполняемые любой сколь угодно сложной детерминированной машиной. С другой стороны, еще никому не удалось описать детерминированную машину, которая бы по широте диапазона производимых ею операций превосходила машины Тьюринга, и маловероятно, что такую машину когда-нибудь удастся найти. Следовательно, мы можем с полным основанием считать, что детерминированным машинам доступны лишь те операции, которые доступны машинам Тьюринга. Попытаемся установить все типы машин Тьюринга и *перенумеровать* их.

Любая машина полностью известна, если мы знаем, как она реагирует на любую комбинацию $S_1 0$, $S_1 1$, $S_2 0$, $S_2 1$, ..., $S_N 0$, $S_N 1$, то есть как машина будет вести себя, когда, находясь во внутреннем состоянии S_i, она увидит 0 или 1. Условимся записывать каждую *реакцию* так, чтобы обозначения внутренних состояний S_i шли в конце. Сами реакции упорядочим следующим образом: реакция на $S_1 0$, реакция на $S_1 1$, реакция на $S_2 0$, реакция на $S_2 1$ и т. д. Выписанные в таком порядке реакции назовем *характеристикой* машины. Пробелы нам не нужны: различные индексы у S_i указывают лишь на то, где заканчивается одна и начинается другая реакция.

Характеристики машин перенумеруем при помощи специального *кода*. Сопоставим символам

$$0,\ 1,\ +,\ -,\ S_1,\ S_2,\ S_3,\ S_4,\ \ldots$$

их кодированные обозначения

$$0,\ 00,\ 01,\ 10,\ 11,\ 000,\ 001,\ 010,\ \ldots,$$

для которых условимся оставлять места с четными номерами. На местах с нечетными номерами в общем случае будут стоять нули. Чтобы отделить друг от друга два смежных символа, вместо пробела будем на четных местах вписывать перед стоящими там символами единицу. Пользуясь кодом, можно описывать различные реакции. Окончание реакции нетрудно определить по появлению S_i.

Кодированное описание машины из примера I имеет, например, следующий вид:

$$1001100011011001101101.$$

Эту последовательность нулей и единиц можно прочитать как число, записанное в двоичной системе (в десятичной системе оно соответствовало бы числу 2 504 681). Таков *кодовый номер* машины из примера I. После кодового номера, отделив его тремя единицами (комбинацией единиц, никогда не встречающейся в записи кодового номера машины), запишем последовательность цифр, стоящую на бумажной ленте (исходные данные). В результате мы получим кодовый номер, позволяющий упорядочить комбинации «машина + лента с исходными данными». Для машины, производящей бесконечную последовательность

$$010101\ \ldots,$$

кодовый номер в каталоге машин, снабженных лентой с исходными данными (начальные данные плюс программа), был бы равен

$$10011000110110011011011110,$$

для машины, производящей бесконечную последовательность

$$101010\ \ldots,$$

кодовый номер был бы

$$10011000110110011011011111.$$

Итак, все комбинации «машина + лента с исходными данными» получили бы свой кодовый номер. По кодовому номеру нетрудно узнать характеристику машины, а также восстановить ее описание («построить машину») и запись на ленте с исходными данными.

Однако это отнюдь не означает, будто любое натуральное число, записанное в двоичной системе, встречается среди кодовых номеров комбинаций «машина + + лента с исходными данными». Даже если какое-то натуральное число, записанное в двоичной системе, встречается среди кодовых номеров в каталоге комбинаций «машина + лента с исходными данными», то эта комбинация не обязательно должна производить бесконечную числовую последовательность. В том случае, когда двоичная запись натурального числа совпадает с кодовым номером комбинации «машина + + лента с исходными данными», производящей бесконечную числовую последовательность, его называют кодовым номером этой последовательности. Так *каждой бесконечной последовательности нулей и единиц, которую может породить машина, мы сопоставляем конечный кодовый номер.* Более того, одна и та же последовательность может иметь даже несколько конечных кодовых номеров, поскольку одну и ту же бесконечную последовательность могут производить различные машины, заправленные лентами с исходными данными.

Если кодовый номер производимой машиной бесконечной последовательности задан, то найти последовательность совсем нетрудно. Отыскав в кодовом номере метку 111, разделим его на кодовый номер машины и запись исходных данных на бумажной ленте. Зная кодовый номер машины, восстановим ее характеристику и тем самым «построим» саму машину. Наконец, введем в нее ленту с исходными данными, и машина заработает.

Представим себе, что перед нами каталог всех производимых машинами последовательностей нулей и единиц. Упорядочим эти последовательности просто по их кодовому номеру. Некоторые из последовательностей будут встречаться в каталоге не один, а несколько раз, но для наших целей это несущественно. Главное, что наш каталог полный: ни одна производимая машинами бесконечная последовательность в нем не пропущена.

Пользуясь терминологией гл. 1, можно сказать, что множество производимых машиной бесконечных числовых последовательностей *счетно*. Но каждому математику известно, что мощность множества всех бесконечных числовых последовательностей более мощна, чем мощность счетного множества. Следовательно, существуют такие последовательности из нулей и единиц, которые не может произвести любая машина, какие бы исходные данные мы ни записывали на ее ленте. Доказывать эту математическую теорему в общем виде не нужно: мы просто построим пример такой «немашинной» последовательности.

Последовательность, которую не могут построить машины

Некоторые натуральные числа N служат кодовыми номерами производимых машинами последовательностей нулей и единиц. Пусть C_N — последовательность с кодовым номером N. Если натуральное число N не совпадает с кодовым номером ни одной из производимых машинами последовательностей, то C_N также можно придать смысл: условимся считать, что в этом случае C_N представляет последовательность, состоящую из одних лишь нулей (попутно заметим, что такая последовательность также принадлежит к числу последовательностей, производимых машинами). Запишем последовательности $C_1, C_2, \ldots$ одну под другой:

$$
\begin{array}{l}
C_1 : a_1^{(1)},\ a_2^{(1)},\ a_3^{(1)},\ \ldots,\ a_p^{(1)},\ \ldots \\
C_2 : a_1^{(2)},\ a_2^{(2)},\ a_3^{(2)},\ \ldots,\ a_p^{(2)},\ \ldots \\
C_3 : a_1^{(3)},\ a_2^{(3)},\ a_3^{(3)},\ \ldots,\ a_p^{(3)},\ \ldots \\
\vdots \qquad \vdots \qquad \vdots \qquad \vdots \qquad\quad \vdots \\
C_p : a_1^{(p)},\ a_2^{(p)},\ a_3^{(p)},\ \ldots,\ a_p^{(p)},\ \ldots \\
\vdots \qquad \vdots \qquad \vdots \qquad \vdots \qquad\quad \vdots
\end{array}
$$

Величины $a_i^{(j)}$ здесь означают нули и единицы. Каждая строка представляет собой производимую машинами последовательность. Любая из производимых машиной последовательностей совпадает с какой-то строкой.

Рассмотрим диагональ нашей бесконечной таблицы, идущую из ее левого верхнего угла в правый нижний угол. Заменим все стоящие на диагонали нули единицами, а все единицы — нулями. Мы получим последовательность

$$D : d_1,\ d_2,\ d_3,\ \ldots,\ d_p,$$

$$\text{где } d_i = \begin{cases} 1, & \text{если } a_i^{(i)} = 0, \\ 0, & \text{если } a_i^{(i)} = 1. \end{cases}$$

Последовательность D заведомо не совпадает ни с одной из строк таблицы. Действительно, последовательность D не совпадает с последовательностью C_1, поскольку у них различны первые элементы, с последовательностью C_2 — поскольку заведомо отличается от C_2 вторым членом,. . ., с последовательностью C_p — поскольку отличается от C_p по крайней мере p-м членом.

Итак, последовательность D не совпадает ни с одной из строк выписанной нами таблицы, а поскольку таблица содержит все производимые машинами числовые последовательности, то последовательность D машины произвести не могут. Тем не менее последовательность D не является случайной, возникающей, например, при записи исходов бесконечной серии бросаний монеты: D — вполне «детерминированная» (неслучайная) последовательность.

Построенный нами пример последовательности, которую не может построить никакая машина, — результат весьма странный, гораздо более странный, чем можно было бы думать.

Универсальные машины

Доказанную нами теорему мы сочли странной потому, что даже *универсальные машины*, способные (если ввести в них ленту с соответствующим образом подобранными исходными данными) построить любую производимую машинами последовательность, не могут произвести последовательность D. Не будем пытаться выписывать в явном виде характеристику универсальной машины. Таких попыток не предпринимал никто, и мы лишь напрасно потратили бы время. Но принципиальное устройство универсальной машины можно описать без особых трудностей.

Если универсальная машина должна выписать производимую машинами числовую последовательность с кодовым номером N, то этот кодовый номер необходимо нанести на ленту с исходными данными. Универсальная машина, найдя метку 111, разбивает двоичную запись числа N на кодовый номер неуниверсальной (специальной) машины и запись исходных данных на ленте, вводимой в специальную машину, — условимся называть эту ленту вторичной. Осмотрев первую клетку на вторичной ленте, универсальная машина обнаруживает в ней либо нуль, либо единицу. Затем она расшифровывает первый отрезок кодового номера специальной машины и устанавливает, каким образом та реагирует на $S_1 0$ или $S_1 1$. Выполнив эти операции, универсальная машина узнает, какую клетку вторичной ленты должна рассматривать специальная машина после первой реакции и в каком внутреннем состоянии S_i та находится. Универсальная машина расшифровывает еще один отрезок кодового номера специальной машины, чтобы узнать, каким образом специальная машина реагирует на $S_i 0$ или $S_i 1$. Продолжая последовательно, отрезок за отрезком, расшифровывать кодовый номер специальной машины, универсальная машина построит числовую последовательность с кодовым номером N.

Что же мешает универсальной машине построить выписанную нами выше числовую последовательность D? Ведь для этого ей нужно лишь строить одну за другой последовательности C_1, C_2, ... (причем каждую последовательность — не полностью, а лишь до диагонального члена), а затем заменить каждый член диагональной последовательности $a_1^{(1)}$, $a_2^{(2)}$, $a_3^{(3)}$, ... на противоположное число, то есть нуль — на единицу и единицу — на нуль. Операции, необходимые для построения последовательности D, совсем просты и доступны «разумению» универсальной машины, если снабдить ее соответствующими инструкциями, записанными на ленте с исходными данными.

И тем не менее мы доказали, что последовательность D невозможно построить при помощи машины. Не противоречит ли одно другому?

Нет, никакого противоречия между доказанным нами утверждением и кажущейся простотой построения последовательности D не возникает. Инструкции, якобы позволяющие универсальной машине построить все члены диагональной последовательности $a_1^{(1)}$, $a_2^{(2)}$, $a_3^{(3)}$, ..., совсем не просты.

Универсальная машина строит последовательность C_p. Машину интересует *окончательное* значение члена $a_p^{(p)}$, стоящего на p-м месте в последовательности C_p, то есть значение, которое принимает величина $a_p^{(p)}$ после того, как машина закончит последнюю операцию с числом, стоящим на p-м месте в последовательности C_p. Но p-е место «не успокаивается» довольно долго: машина то что-то вписывает, то что-то вычеркивает, и значения, принимаемые $a_p^{(p)}$, никак не «стабилизируются». Каким образом универсальная машина узнает, принял ли p-й член $a_p^{(p)}$ последовательности C_p окончательное значение и с какого момента это произошло? Предположим, что на p-м месте довольно долго стоит нуль или единица. Каким образом машина определит, сохранится ли число, стоящее на p-м месте, навсегда или еще изменит свое значение? Если p — кодовый номер некоторой производимой машинами последовательности, то, начиная с некоторого момента, диагональный член $a_p^{(p)}$ «успокаивается» — принимает свое окончательное значение. Но когда именно наступает успокоение? Ответ на этот вопрос необходим для того, чтобы машина могла установить значение $a_p^{(p)}$ и, определив тем самым d_p, перейти к $p+1$. Если p не является порядковым номером (в первоначальном смысле) ни одной из производимых машинами последовательностей, то $a_p^{(p)}$ считается равным нулю. Дойдя до очередного числа p, машина должна уметь находить ответ на вопрос: совпадает оно с кодовым номером одной из производимых машинами последовательностей или нет? Для этого в машину с самого начала должна быть введена программа, позволяющая ей находить ответ на этот вопрос. Машина, умеющая отвечать на перечисленные вопросы, могла бы выписать последовательность D. Мы же только что доказали обратное: никакая машина не может произвести последовательность D.

Что отсюда следует? А то, что нельзя построить универсальную машину, которая могла бы разрешать все подобные вопросы. Как бы мы ни конструировали универсальную машину, вопрос о том, чему равен элемент $a_p^{(p)}$ последовательности C_p — нулю или единице, останется для нее неразрешимым, сколько бы она ни работала. Впрочем, вполне может случиться, что человек, наблюдающий за работой машины извне, путем каких-то остроумных умозаключений сможет разрешить вопрос о том, чему равен член $a_p^{(p)}$ последовательности C_p. Мы утверждаем лишь, что этот вопрос неразрешим для машин. Не исключено, что вопрос вполне разрешим другими, не машинными средствами.

Итак, начав с таких конкретных вещей, как вычислительные машины, мы дошли до весьма абстрактных рассуждений. Но это еще не все.

Некогда думали, что можно построить такую машину, которая сможет решать математические задачи следующего типа: останется ли навсегда при построении бесконечной последовательности C_p на месте p-го члена число 1 или число 0? Такие вопросы нельзя решать при помощи машин, ответы на них необходимо искать в строгих математических доказательствах.

Математическое доказательство состоит из конечной последовательности высказываний, каждое из которых является следствием предыдущего. При доказательстве очередного высказывания математик имеет право воспользоваться конечным числом правил, называемых правилами вывода. Эти правила вполне формальны и позволяют по внешнему виду двух высказываний узнавать, выводимо ли из них третье высказывание.

Предположим, что нам удалось найти код, позволяющий записать любое математическое выражение, любую формулу и любое высказывание в виде набора нулей и единиц. Тогда каждое математическое выражение, каждая формула и каждое высказывание получат свой кодовый номер, по которому без труда можно восстановить скрытое за ним математическое выражение, формулу или высказывание. Каждое доказательство также получает свой кодовый номер, составленный из кодовых номеров входящих в него высказываний так же, как ранее мы составляли кодовый номер машины из кодовых обозначений ее характеристики и исходных данных. Взяв любое натуральное число, попытаемся выяснить, не совпадает ли оно с кодовым номером некоторого математического выражения, формулы, высказывания или доказательства. Для того чтобы машина могла заниматься подобной математикой, нам следует объяснить ей, каким образом она должна определять, что означает (и означает ли что-нибудь) заданное число. Машина может сформулировать высказывание «Число x совпадает с кодовым номером некоторого утверждения» и затем проверить его истинность при любом значении x. Не составляет труда проверка истинности при любом значении x и таких высказываний, как, например, «Число x совпадает с номером некоторого высказывания, содержащего одну переменную» или «Высказывание с кодовым номером доказуемо». Но машина может отпечатать высказывание Q, гласящее: «Высказывание с кодовым номером x не доказуемо» (непосредственно проверить его истинность при любом заданном x машина просто не в состоянии, поскольку для этого ей пришлось бы перебрать все доказательства и выяснить, не позволяют ли они установить истинность высказывания с номером x).

Рассмотрим высказывание Q подробнее. Пусть q — его кодовый номер. Высказывание Q содержит одну переменную x. Если вместо x мы подставим q, то получится высказывание, не содержащее ни одной переменной. Обозначим его P. Спрашивается, доказуемо ли высказывание P?

Высказывание P возникает при подстановке в Q вместо переменной x кодового номера q высказывания Q. Следовательно, высказывание P эквивалентно следующему: «Не доказуемо, что высказывание с кодовым номером q не доказуемо». Таким образом, высказывание P содержит утверждение о своей собственной недоказуемости. Ясно, что такое высказывание не доказуемо. Действительно, если бы оно было доказуемо, то мы могли бы доказать его недоказуемость, что противоречит самому высказыванию. Следовательно, высказывание не доказуемо, хотя и опровергнуть его также невозможно: к выводу о недоказуемости P мы пришли путем логических рассуждений.

Итак, мы обнаружили высказывание, которое машина не может ни доказать, ни опровергнуть. Если бы машина стала доказывать одно за другим все возможные высказывания, то среди них никогда бы не встретилось ни высказывание P, ни его отрицание.

Впрочем, мы можем утверждать это лишь с известными оговорками. К выводу о недоказуемости высказывания P мы пришли, рассуждая от противного: иначе высказывание и его отрицание были бы доказуемы одновременно. Вывод о существовании неразрешимых для машин высказываний был получен в предположении, что машина не может доказывать противоречивые высказывания.

Это знаменитая теорема К. Гёделя (1931) о неразрешимости. Мы пытались весьма сжато наметить схему ее доказательства. После столь напряженных усилий обратимся к менее абстрактной области и посмотрим, как там решается вопрос о возможностях вычислительных машин.

Шахматные автоматы

Нам уже приходилось строить машину для игры в ним. Сделать это было довольно просто. Пробовали мы обучать машины и азартным играм. Но существуют еще и такие игры, как шахматы. Их нельзя отнести к числу азартных игр, хотя азарт не чужд шахматистам. Они вполне могут совершать ошибочные ходы сами и использовать ошибки, совершенные их партнерами. Для беспроигрышной игры в ним требуется знать лишь «формулу успеха» и уметь правильно считать. В шахматах такой формулы не существует. Взглянув на начальную позицию на шахматной доске, невозможно предсказать, кто выигрывает — черные или белые — и закончится ли при правильной игре партия вничью. Правда, существует так называемая шахматная теория, но она целиком эмпирична. Любая шахматная партия всегда представляет собой новую, еще не решенную задачу, поэтому играть в шахматы несравненно интереснее, чем в ним, где все заранее известно.

Идея создания шахматных автоматов очень стара. О таких автоматах мечтали еще философы XVII в.

А уже в XVIII в. появились изобретатели, начавшие строить шахматные автоматы, причем один причудливее другого. Наибольшей славой пользовался автомат в виде турка в чалме, курящего кальян. Играл автомат довольно сильно и одерживал победу почти над всеми живыми партнерами, с которыми ему доводилось встречаться. Электронный мозг в то время еще не был изобретен, и для игры в шахматы был необходим «человеческий элемент». И действительно, среди нагромождения псевдоаппаратуры в автомате прятался человек миниатюрного сложения, делавший за турка все ходы.

Электронный мозг, умеющий играть в шахматы, не создан и поныне, и вряд ли стоит затрачивать силы и средства на конструирование «машины-шахматиста». Однако задуматься над тем, каким должно быть устройство такой машины, стоит. Разумеется, говоря об устройстве, мы имеем в виду внутреннее устройство машины. Внешне оно выглядело бы, как большой шкаф. В XVIII в. шахматному автомату приходилось придавать человеческое обличье. Ныне никому не приходит в голову требовать, чтобы автомат, продающий почтовые марки, по внешнему виду походил на почтового служащего.

Прежде всего шахматная машина должна была бы содержать какие-то ячейки (реле, электронные лампы, полупроводниковые диоды), соответствующие 64 полям шахматной доски и 32 фигурам. Если на каком-то поле стоит фигура, то ячейки, соответствующие этому полю и фигуре, соединяются. Одновременно открываются «двери», ведущие к полям, которые данная фигура держит под ударом. Если этой фигуре преграждает путь другая фигура, то двери, ведущие к полям, до которых данная фигура не может добраться, закрываются. Разумеется, машина должна быть осведомлена обо всех правилах игры в шахматы, о рокировках, о взятии фигур на проходе, о мате и пате, о ничьих при вечных шахах и тому подобном.

Нужно ли удивляться, что машина, способная вместить всю шахматную премудрость, имела бы внушительные размеры и весьма сложное внутреннее устройство. Заглянув под крышку шахматного автомата, мы увидели бы кажущийся хаос клемм и проводов, в действительности смонтированных в тщательно продуманном порядке. Но не о трудностях монтажа пойдет сейчас речь — с этой чисто практической задачей справится любой техник.

Предположим, что все трудности преодолены и перед нами машина, умеющая играть в шахматы. Это означает, что машина знает все правила игры в шахматы, умеет их применять и протестует, когда ее партнер нарушает правила. Во всем остальном она уступает любому человеку, едва выучившему правила игры в шахматы и после непродолжительной тренировки решившему сыграть свою первую партию. Прежде чем сделать очередной ход, начинающий шахматист может поразмыслить, и, сколь бы неуклюжими ни были его ходы, все же он подготовлен к партии лучше, чем машина, которая вообще лишена органа мышления и в лучшем случае способна выбирать ходы наугад.

Бессмысленно винить машину, если в игре она уступает даже начинающему шахматисту: мы сами должны были бы позаботиться об устройстве, позволяющем ей на любой ход противника отвечать наилучшим из возможных ходов. Именно так мы поступили, конструируя машину для игры в ним. Именно так нужно поступить и с шахматным автоматом.

Наша задача будет выполнена, если нам удастся разработать релейно-контактную схему, позволяющую выбирать наилучший в данной позиции или по крайней мере не слишком плохой ход. При игре в ним все обстояло очень просто: хороший ход определялся одним арифметическим правилом, и конструктору оставалось только перевести это правило на язык релейно-контактных схем.

Совершенно иная ситуация возникает при анализе игры в шахматы. Теория игры в ним полностью известна, шахматной же теории просто не существует. То, что принято называть шахматной теорией, представляет собой в действительности свод тактических приемов и традиций, быть может ошибочных от начала и до конца. Еще на заре XX в. полагали, что первый ход следует делать королевской пешкой. Ныне стремятся открыть выход ферзю. Вполне возможно, что оба правила неверны и истина заключается в чем-то ином, по-настоящему еще никем не открытом и настолько противоречащем привычным представлениям, что того, кто осмелится когда-нибудь играть по-новому, сочтут безумцем.

Итак, шахматной теории не существует, хотя мы и намеревались ввести ее в машину. Разумеется, нашим планам не суждено сбыться. Тем не менее люди предпочитают играть в шахматы, не имеющие теории, а не в ним с его полной законченной теорией. Более того, именно отсутствие теории придает шахматам особую привлекательность.

Шахматной машине не остается ничего другого, как подражать шахматисту-человеку: она должна обдумывать свои ходы. Как это делать? Так же, как это делает человек. А как он это делает?

«Он думает» — такой ответ вряд ли удовлетворит нас, поскольку не позволяет ни на шаг продвинуться в обучении электронного шахматиста. Гораздо больше полезной информации мы смогли бы извлечь, если бы ответ гласил: «Размышляя, человек продолжает шахматную партию, но при этом не притрагивается к фигурам. Он играет на воображаемой шахматной доске, где может взять назад любой показавшийся неудачным ход. При этом он играет не только за себя, но и за своего противника, имитируя его ходы. Человек анализирует несколько вариантов и останавливается на том, который по его представлениям сулит наибольшие преимущества. Лишь после этого человек совершает на реальной доске первый ход из отобранного варианта».

Так же должна поступать и машина: проигрывать внутри себя все варианты. Предположим, что в любой позиции машина оценивает десять возможных продолжений и каждый вариант просматривает на десять ходов: пять своих и пять противника. Тогда общее число различных ходов равно $10^{10} = 10\,000\,000\,000$. Если на один ход мы дадим ей одну стотысячную долю секунды, то это означает, что над очередным ходом машина будет думать целый день. Срок, безусловно, слишком долгий.

Мы предположили, что в любой позиции машина оценивает 10 возможных продолжений. Нельзя ли довольствоваться меньшим? В действительности число возможных продолжений гораздо больше, и, чтобы выбрать из них лишь 10, машина должна обладать неким критерием. Разумеется, критерий отбора должен храниться втайне от партнера, с которым играет машина, в противном случае партнер может легко поймать ее в какую-нибудь ловушку. Но существует еще одна трудность. Машина должна записывать анализируемые ею варианты, ибо в противном случае она не смогла бы сравнивать их, прежде чем остановить свой выбор на лучшем (с ее точки зрения). На магнитной ленте можно записать примерно 100 знаков на квадратный сантиметр. Если для записи каждого хода требуется затратить 10 знаков, то для записи просматриваемых машиной 10^{10} ходов понадобится магнитная лента длиной 1 км и шириной 100 м. Оперировать с такой «памятью» было бы невозможно.

Но это еще не самая большая трудность. Каким образом может машина, анализируя 10^{10} возможных ходов, определять лучшие и худшие из них? Какими принципами она руководствуется при отборе? Как устанавливает, что тот или иной ход наилучший? Все обстоит очень просто, когда какие-то из просматриваемых вариантов приводят к мату, но так бывает лишь в редких случаях. Выигрыш фигуры или качества гораздо менее убедителен. Кроме того, если противнику известно, что машина, начиная с пятого хода, живет надеждой на взятие фигуры, он может, разыграв гамбит, без особого труда обмануть ее.

Итак, мы пришли к неутешительному выводу: нам не удастся построить машину, которая бы действительно играла в шахматы. И все же мы знаем, что человек играет в шахматы, хотя неизвестно, как он это делает. Проведенный нами «анализ» действий шахматиста при игре в шахматы далеко не полон. Мы чем-то пренебрегли. Вопрос в том, чем же?

Почему человеку не приходится перебирать 10^{10} различных ходов, прежде чем он сделает очередной ход? Откуда человек получает недостающую машине информацию? Шахматист за время шахматных баталий накапливает опыт и хранит его в своей памяти. Все ясно: машине также нужна память, которая накапливала бы весь нужный опыт и давала бы совет в случае необходимости.

Человеческий мозг состоит примерно из 10^{10} клеток, называемых нейронами. Не приходится и мечтать о том, чтобы создать нечто подобное средствами современной техники. Для размещения 10^{10} ячеек памяти необходима, как мы подсчитали, магнитная поверхность размерами 1000×100 м, что совершенно нереально. Клетки человеческого мозга гораздо меньше электронных ячеек памяти. Нейроны оперируют не с электронными лампами и полупроводниками, а с молекулами — молекулой протеина, прикрепленной в определенном месте к молекуле рибонуклеиновой кислоты.

Создатели миниатюрных ячеек и компактной машинной памяти одерживают один за другим все более крупные успехи. И все же в этой области технике еще долго не удастся догнать природу. Впрочем, нет никакой необходимости копировать человеческий мозг целиком. Преследуются более скромные цели построить действующую модель лишь той части мозга, которой человек пользуется, играя в шахматы. Чтобы читатель мог уяснить подлинные масштабы отставания искусственного мозга от его естественного прототипа, упомянем лишь о том, что объем памяти большой современной вычислительной машины примерно равен объему памяти дождевого червя. Для игры в шахматы этого явно недостаточно.

Но не будем падать духом! Предположим, что проблема создания компактной памяти решена и машина обладает памятью, объем которой сравним с памятью человека. Машина помнит все сыгранные ею партии и все анализы, проведенные ею во время этих партий и на досуге. Для этого машине потребовалась бы колоссальная память, но мы согласны пойти на все и предоставили ей такое гигантское хранилище информации. Встретив любую позицию на шахматной доске, машина начинает рыться в своей памяти и, если ей удается найти такую же позицию, читает свои прежние записи и лишь после этого совершает очередной ход.

Быть может, кому-нибудь из читателей покажется, что уж теперь-то проблема создания шахматной машины решена. Увы! До желанной цели по-прежнему еще очень далеко. Но кое-чего мы все же добились:

перед нами начала вырисовываться истинная трудность. Гигантская память, которой мы наделили машину в своем воображении, оказалась излишней. Машина не извлекает ни малейшей пользы из того, что может устанавливать полную тождественность позиции, в которой ей предстоит сделать очередной ход, с позицией, уже встречавшейся ей ранее. Если отвлечься от позиций, возникающих в традиционных дебютах, то вероятность встретить за всю шахматную жизнь две одинаковые позиции чрезвычайно мала. Накопленный машиной опыт, все партии, которые она столь тщательно записывает в своей памяти, понадобятся ей крайне редко.

Машина недостаточно хорошо подражала человеку-шахматисту. Он никогда не хранит в своей памяти сыгранные им партии целиком. Шахматист отбирает в каждой партии наиболее существенное и опускает второстепенные детали. Искусство запоминать включает в себя и искусство забывать.

Приведем пример. Начинающему шахматисту, королю которого на c7 и ладье на e8 в одной из сыгранных им партий одновременно угрожал конь противника на a8, нелегко забыть последовавшую потерю качества, даже когда прочие детали позиции, сложившейся на шахматной доске, изгладятся из его памяти. Начинающий шахматист извлек урок из постигшей его неудачи и знает, как парировать аналогичную угрозу, если она возникнет в дальнейшем. Наш шахматный автомат незнаком с понятием «аналогичная угроза». Полное совпадение или несовпадение позиций — вот все, что он способен воспринимать. Если в одной из последующих партий конь противника снова будет угрожать двум фигурам нашего шахматиста, он сумеет избавиться от «вилки», даже если на шахматной доске сложится совсем иная позиция, чем в первый раз. Машина может что-то извлечь из накопленного ею опыта лишь тогда, когда на шахматной доске сложится позиция, до мельчайших деталей совпадающая с встречавшейся ранее. Использование неполного совпадения, аналогии, абстракции, умение отличать существенное от второстепенного составляет отличительные особенности человеческого мышления. Все это мы упустили из виду, анализируя игру шахматиста - человека. Современные автоматы сконструированы так, что они могут распознавать лишь полное тождество. В этом их сила и слабость. Автомат, продающий сигареты, настолько слабоумен, что объявит «забастовку», если вместо мелкой монеты в него опустить крупную банкноту. Необходимо было бы создать автоматы совершенно иного типа, но для этого сначала мы должны понять, что, собственно, означает подобие, аналогия, абстракция, существенное, второстепенное и т. д.

В некоторых магазинах изучают статистику покупателей и подсчитывают число входящих и выходящих посетителей. (Кстати сказать, неизвестно почему, но входящих оказывается больше.) Кое-где людей, занимающихся таким подсчетом, заменили электронными счетчиками, прощупывающими покупателей невидимыми лучами. В небольших кафе и банках часто можно видеть швейцара, предупредительно открывающего перед посетителями входную дверь. В его обязанности входит не только вести счет посетителям, но и приветствовать их. Если швейцар вежлив, то при виде дамы он непременно добавит к обычному «Добрый день» учтивое «мадам». Мне неоднократно приходилось бывать в подобных кафе, и, должен сказать, я ни разу не видел, чтобы швейцар, действующий почти автоматически, ошибся. (То, что швейцар не делает различия между почтенной дамой и юной девушкой, для нас не столь существенно.)

Не исключено, что наши критерии столь же просты, как критерии, которыми пользуются птицы и звери. Свою догадку мы могли бы проверить экспериментально, показывая подопытному фотографии различных частей человеческого тела и опрашивая, кому они принадлежат — мужчине или женщине. Действуя таким образом, нам удалось бы обнаружить важнейшие критерии. Если их окажется слишком много, то надеяться на то, что найденные критерии удастся ввести в машину, не приходится. В этом случае мы будем вынуждены прибегнуть к другому методу. Расставив перед машиной группу людей, мы объясним машине, кто из них мужчина и кто женщина. Коллекцию фотопортретов, надлежащим образом типизированных и переведенных на язык машины, введем в память автомата. В случае необходимости машина может извлечь соответствующий портрет из архива и, взглянув на него, сравнить с новым лицом. Разумеется, машина должна уметь устанавливать не только полное тождество сравниваемых портретов, но и их сходство, то есть абстрагироваться от несущественных черт и выделять главные. Для этого машину необходимо научить отличать мужчин и женщин, причем «образование», возможно, придется продолжить и после того, как автомат займет место швейцара у дверей кафе. Если машина ошибется, ее условно накажут, и это событие навсегда останется в ее памяти. Так, постепенно совершенствуя способности машины, мы в конце концов получим автомат, умеющий безошибочно, с первого взгляда отличать мужчин от женщин. Но самое главное при этом будет то, что мы научим машину вырабатывать абстрактные понятия.

Образование абстракций машинным интеллектом — важная философская проблема, имеющая практическое значение. Во многих конструкторских бюро уже сейчас разрабатываются проекты машин, умеющих читать печатный и даже рукописный тексты. Если машина разбирает мой почерк и может так переработать текст, что другая машина возьмется перепечатать его, то отпадает необходимость в услугах стенографистки-машинистки. Каким образом секретарю удается расшифровывать мои записки, если ни одна из них не похожа на другую? Пока еще не существует машины, способной проделать то, что ежедневно с легкостью делает секретарь. А такие машины очень нужны — машины, умеющие читать чеки, адреса, телефонные справочники и тому подобную «литературу». Конструкторам никак не удается научить машины читать тексты, напечатанные на пишущей машинке или типографским способом, поскольку между буквами, которые кажутся нам одинаковыми, в действительности имеются небольшие различия. Одинаковыми буквы представляются нам лишь потому, что мы обладаем способностью абстрагироваться от несущественных деталей, хотя и не могли бы объяснить, что существенно и что второстепенно.

Высшее достижение наших конструкторов в настоящее время — машина, автоматически печатающая текст, который задан ей на перфокартах или перфоленте. Единственное, что требуется сообщить ей дополнительно, — правила переноса. Однако даже это оказывается достаточно сложной задачей. Машины, умеющие переводить с одного языка на другой, находятся еще в стадии изучения, но над созданием электронного переводчика трудится множество людей различных специальностей от лингвистов до математиков.

Создание машины, способной отличать мужчин от женщин, отнюдь не является актуальной проблемой. Сконструировать такую машину было бы сравнительно просто. Построить шахматную машину несравненно труднее. Ее необходимо научить абстрактному мышлению, но никто не может сказать, как происходит процесс образования абстракции, хотя каждый знает, что означает это слово. Машина ставит перед нами сложную философскую проблему, поскольку она охотно научилась бы образовывать абстракции (как, впрочем, и многому другому).

Но представим себе, что проблема решена, все трудности преодолены и построен автомат, способный распознавать и абстрагировать от конкретных деталей позиции поля, находящиеся под ударом коня, а также образовывать другие абстрактные понятия, которые необходимы сильному шахматисту. Назовем такой автомат Каисс. До того, как Каисс сыграет свою первую партию, он не более чем неопытный новичок в шахматах. Играя с Каиссом в шахматы, мы обучаем его. Если Каисс играет хорошо, то получает от нас поощрение. Если он делает неверный ход, то мы его наказываем. Каисс очень прилежный ученик. Стоит нам оставить его одного, как он погружается в анализ сыгранных партий. Времени у него много: он может целиком посвятить себя шахматам. Каисс не спит, не умывается, не чистит зубы (их у него нет), ему не нужно бриться. Питается Каисс электрическим током, который поступает к нему бесперебойно. О нем заботится специальный служитель, стирающий пыль с его поверхности. Если Каисс заболеет, к его услугам врач — специалист по электронике. Довольно скоро Каисс по силе игры превзойдет своего учителя, поскольку тот может уделять шахматам лишь часть своего времени. Сверхчеловеческому гению Каисса не составит труда выиграть шахматную корону мира.

Если шахматное образование Каисса меня особенно не интересует, то я могу поступить иначе: построить два автомата, создать не только Каисса, но и Каиссу. Нехорошо, чтобы Каисс все время оставался в одиночестве. Играя, Каисс и Каисса будут обучать друг друга так же, как обучают друг друга в совместных играх близнецы. Им известен лишь один вид деятельности, ведома лишь одна страсть — шахматы. Конструкторы не вложили в Каисса и Каиссу орган ревности, ненависти или любви. В равной мере Каисс с Каиссой гарантированы от неожиданного появления у них нежелательного потомства.

Когда человек играет в шахматы, он размышляет. Могут ли мыслить Каисс и Каисса?

Могут ли машины мыслить?

Если под «мышлением» понимать всю разнообразную умственную деятельность человека, то на поставленный вопрос придется дать отрицательный ответ. Каисс и Каисса — чрезвычайно узкие специалисты, их ограниченность во много раз превосходит ограниченность самых узких специалистов среди людей.

Какие бы машины мы ни конструировали, они всегда будут специалистами. Именно узкая специализация в какой-то одной области позволяет машинам развивать деятельность, превосходящую человеческие возможности. Если требуется создать нечто достойное называться человеком в полном смысле этого слова, то вовсе не требуется обращаться к конструкторам дорогостоящих машин. Метод, дарованный природой, по-прежнему остается самым дешевым и в наши дни, когда почти все производство находится в руках специалистов, пользуется тем же всеобщим признанием, как и во времена Адама и Евы.

Но если ограничиться рассмотрением игры в шахматы, то все же можно спросить, можно ли назвать то, что делают во время игры Каисс и Каисса, мышлением. «А почему бы и нет? — ответил бы я. — Каисс и Каисса — сильные шахматисты и выигрывают свои партии не чудом, а при помощи точного анализа, мало чем отличающегося от анализа, проводимого шахматистами-людьми».

Такой вывод многим не нравится. Они склонны считать, что мышление сопровождается неким таинственным процессом. Быть может, это и так, но не менее вероятно и то, что подобные взгляды ошибочны. Когда думает кто-то другой, я воспринимаю лишь сообщаемый им результат (иногда лишь удается заметить на лице собеседника мимическую игру). Мыслящие машины выдают нам результат, который по качеству неотличим или даже превосходит продукт мышления человека. Если кто-нибудь не может обойтись без внешних симптомов мышления, то не составляет труда придать машине облик Мыслителя Родена, наделить ее лбом, который она могла бы морщить, или плечами, чтобы она могла пожимать ими в зависимости от того, что именно требуется в заданный момент. Ясно, что такие требования смешны.

А как обстоит дело с самосознанием машины? Этого я не знаю, поскольку не имею ни малейшего представления о том, что такое самосознание. Можно ли интерпретировать самосознание как память? Если да, то машина в высокой мере наделена самосознанием. Может быть, самосознание надлежит интерпретировать как способность произносить слова: «Я есмь»? Тогда мы вмонтируем в машину магнитофон, который по нескольку раз в день будет произносить магическую формулу «Я есмь». Быть может самосознание означает нечто большее? Весьма правдоподобное предположение, и не исключено, что именно при изучении машин мы столкнемся с необходимостью новой интерпретации самосознания.

Машина сравнима с человеком во многих отношениях. Но человек мыслит целеустремленно, как говорят нам психологи, а машине этого не дано. Мы можем наделить машину способностью мыслить целеустремленно. Машина будет не только действовать последовательно, шаг за шагом, тщательно анализируя все производимые действия, но и мыслить «наоборот», представляя себя в определенном положении в будущем и связывая путем обратных рассуждений свое будущее положение с тем, в котором она находится в данный момент. Это и есть целеустремленность. Или, быть может, целеустремленность означает нечто большее?

Наконец, нельзя не упомянуть о свободе воли — проблеме весьма «колючей», обсуждать которую можно долго и безрезультатно. Что такое свобода воли? В каких случаях мы говорим, что человек свободен в принятии своих решений? Мне кажется, в тех случаях, когда он делает не в точности то, что мы от него ожидаем, когда действия его нельзя предсказать с уверенностью. Это необходимое, но не достаточное условие свободы воли. Если я не могу целиком и полностью положиться на человека, если он никогда не ведет себя так, как хотелось бы мне, если он колеблется, как флюгер на ветру, то я назову его не свободным, а рабом своих страстей. Не он управляет ими, а они всецело правят им. Каков же итог? Человек, обладающий свободой воли, не должен зависеть от моих предсказаний, но не должен и быть независимым от них.

Наша машина не обладает свободой воли. Она точно выполняет то, что мы могли бы (хотя и поздно) предсказать: находит лучший ход в любой позиции. Чтобы обладать свободой воли, она должна была бы в большей степени походить на человека, должна была бы утратить непогрешимую точность своих ходов. Время от времени машине пришлось бы делать слабые ходы или даже нарушать правила игры в шахматы. Иногда машина должна была бы колебаться, прежде чем согласиться на проведение партии, иногда в середине игры — перемешивать фигуры на доске. Все эти действия машине следовало бы совершать не слишком часто и не через правильные промежутки времени. Момент времени, когда машина начинает проявлять свободу воли, должен определять датчик случайных чисел. Достаточно ли всего сказанного для того, чтобы признать за машиной свободу воли, или свобода воли означает нечто большее?

Признаки, по которым мы сравнивали человека и машину, могли показаться поверхностными. Что произошло бы, если бы нам удалось глубже заглянуть в мозг человека? Что происходит в электронном мозгу, известно досконально. Относительно любой его детали специалист может сказать, для чего она служит. Живой мозг по своей сложности во много тысяч раз превосходит электронный. Сложность мозга столь велика, что отдельными тонкостями его строения нам приходится пренебрегать. Многое из того, что мы знаем о функционировании мозга, известно лишь в общих чертах. Мозг состоит примерно из 10^{10} клеток, называемых нейронами. Они образуют огромную «переключательную» схему — нервную систему, позволяющую человеку и животным совершать целеустремленные и осмысленные действия. Переключатели нервной системы — нейроны — так же как реле, электронно-вакуумные трубки или полупроводники, могут находиться в двух состояниях — возбужденном и невозбужденном. Любые два нейрона соединены с третьим так, что он перерабатывает их состояния по схеме одной из логических связок: «и», «или», «если..., то...». Вычислительная машина — несомненно большое чудо. Нервная система — чудо великое. Исходя из того, что мы знаем о меньшем чуде, можно попытаться лучше понять большее.

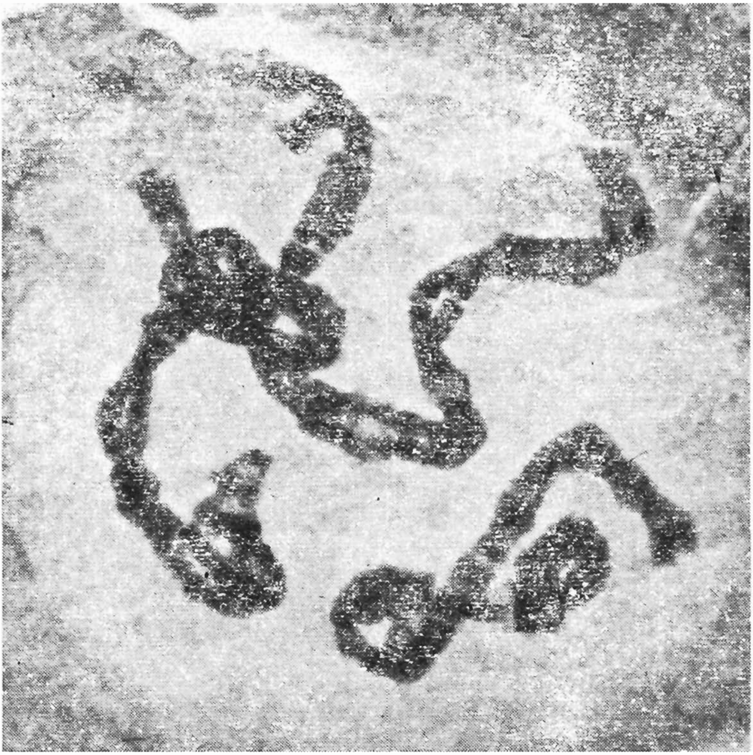
По сравнению с человеческим мозгом объем памяти даже самых больших из современных вычислительных машин невелик. Но машины еще вырастут. Будет ли создана универсальная доказывающая машина, способная доказать все, что может доказать любая другая машина, или шахматный автомат, сравнимый по силе игры с шахматистами-людьми? В это трудно поверить. На пути к созданию таких машин возникают технические проблемы головокружительной сложности. К тому же существует немало гораздо более важных задач.

Стоило ли уделять столько времени перечислению того, что машина может делать лишь в нашем воображении? Но ведь числа и фигуры также существуют лишь в нашем воображении. Люди заинтересовались ими раньше, и числа и фигуры ныне превратались в средства, позволяющие человеку понять и покорить природу. Таких средств много. Одним из них служит вычислительная машина — реальная машина, начисляющая заработную плату и управляющая космическим кораблем, и машина, пленяющая нашу мысль и окрыляющая наше воображение.

5*

Глава 4

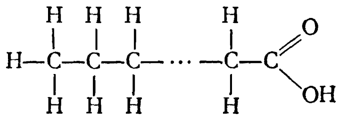
Азбука жизни



Генетический код

Слияние спермия с яйцеклеткой определяет все наследственные признаки нового существа. В крохотном, микроскопически тесном пространстве клетки начертан определенный план. При каждом делении клетки он копируется, чтобы всякая клетка знала, в чем состоит ее задача. В этом строительном плане заранее предусмотрено, какие черты у нового существа будут общими с его собратьями по биологическому виду и какие особые приметы отличат его от многочисленных «родственников». Так, в моем плане записано, что я человек, а не анемон, не амёба, не певчая птица, не пресмыкающееся, не кошка. Там указано, как должны функционировать мои мышцы, печень, нервы, предопределены способность видеть и говорить, цвет кожи, волос и глаз, мужчина я или женщина, левша или «правша», дальтоник или диабетик, — словом, многочисленные подробности моего телосложения, обмена веществ и психического склада.

В основе многообразных проявлений живой природы лежит многообразие форм *протеинов* (белков). По своей химической структуре протеины представляют собой длинные цепи (в действительности даже не одинарные, а двойные) *аминокислот*. Чтобы получить молекулу аминокислоты, достаточно в молекуле какой-нибудь жирной кислоты



заменить атом водорода, не входящий в гидроксильную группу —OH, на аминорадикал $—NH_2$. С той стороны молекулы, где находится ацильная группа —COOH, аминокислота ведет себя как кислота. С той стороны молекулы, где находится аминогруппа $—NH_2$, аминокислота ведет себя как основание. Поэтому аминокислоты можно связывать в длинные цепочки. Так возникают протеины, состоящие из сотен и сотен тысяч аминокислот.

Можно представить себе миллионы различных аминокислот. Двадцати из них выпала особая честь — быть «кирпичиками», из которых построены протеины, одни и те же двадцать аминокислот для всей органической природы от бактерий до *Homo sapiens*. Это счастливое обстоятельство позволяет нам поддерживать свою жизнь, употребляя в пищу других членов великого сообщества живых организмов. Способностью синтезировать аминокислоты под действием света из воздуха обладают лишь растения. Необходимые нам аминокислоты мы получаем от растений либо непосредственно (если мы вегетарианцы), либо косвенным путем (если мы употребляем в пищу мясо животных). Из аминокислот мы должны синтезировать свои протеины. «Кирпичики» различных сортов укладываются в цепочки, отличающиеся друг от друга и по числу кирпичиков, и по порядку, в котором они следуют друг за другом, и образуют молекулы протеинов — как бы двадцатибуквенные «слова». Структура одних протеинов с совсем короткой цепочкой полностью известна. Таков, например, инсулин, играющий важную роль при расщеплении в организме углеводородов. Оказалось, что инсулин одних млекопитающих отличается по своему строению от инсулина других млекопитающих. Протеины с более длинной углеродной цепочкой отличны даже у индивидуумов, принадлежащих к одному и тому же виду.

В строительном плане, содержащемся в оплодотворенной клетке, записано, каким образом новое существо должно синтезировать свои протеины. Наследственный строительный план хранится в так называемых хромосомах (у человека их 46), которые при более близком рассмотрении распадаются на тысячи отрезков. Каждый такой отрезок соответствует какому-то наследственному свойству. Это не что иное, как *дезоксирибонуклеиновые кислоты* (ДНК), структура которых была установлена Уотсоном и Криком в 1952 г. Молекулы ДНК имеют форму спиралей, точнее, двойных спиралей, составленных из небольшого числа «кирпичиков», двух пиримидиновых оснований (тимина и цитозина) и двух пуриновых оснований (аденина и гуанина). Эти четыре основания располагаются в молекуле ДНК в весьма причудливых комбинациях. Каждую молекулу ДНК можно рассматривать как длинное слово, составленное лишь из четырех повторяющихся букв.

Именно рибонуклеиновые кислоты управляют синтезом протеинов. (Условимся в дальнейшем отбрасывать приставку «дезокси». Прежде чем рибонуклеиновые кислоты займутся синтезом протеинов, они подвергнутся еще одной более серьезной трансформации.)

Наследственная информация, закодированная в рибонуклеиновых кислотах, передается дальше двояким образом. Во-первых, она копируется при каждом делении клетки, во-вторых, при синтезе протеина происходит перевод информации с языка, использующего четырехбуквенный алфавит (язык рибонуклеиновых кислот), на язык, использующий двадцатибуквенный алфавит (язык протеинов). Впрочем, встречаются в природе живые существа, знакомые лишь с первым языком: это вирусы, почти целиком состоящие из рибонуклеиновых кислот. Попадая в клетки своего носителя, они находят в них необходимый материал и начинают размножаться.

Для копирования природа придумала весьма хитроумный метод. Рибонуклеиновые кислоты представляют собой двойные спирали, половинки которых относятся друг с другу как негатив и позитив одной и той же фотографии. Против цитозина всегда располагается гуанин и, наоборот, против тимина — аденин. Когда возникает необходимость в снятии точной копии с молекулы рибонуклеиновой кислоты, двойная цепочка распадается по длине, а каждая половинка вновь достраивается до двойной спирали, с позитива печатается негатив, с негатива — позитив.

Второй способ передачи наследственной информации несколько сложнее: он представляет собой перевод с языка рибонуклеиновых кислот, использующего четырехбуквенный алфавит, на язык протеинов, использующий двадцатибуквенный алфавит. Соответствие между тем и другим языком устанавливает некий код. Биохимикам пришлось немало потрудиться, прежде чем удалось его раскрыть.

Возможно ли вообще записать в четырехбуквенном алфавите всю информацию, представимую в алфавите из 20 букв? Оказывается, подобная задача вполне разрешима. Более того, любое слово, например, немецкого языка, использующего обычный латинский алфавит с 26 буквами, можно передать азбукой Морзе, содержащей лишь 2 буквы: точку и тире. Правильнее было бы сказать, что азбука Морзе содержит 3 буквы: точку, тире и пробел, поскольку без пробела мы бы не знали, где начинается и заканчивается комбинация точек и тире, передающая ту или иную букву латинского алфавита.

О коде жизни можно было бы сказать, что это код, не содержащий пробелов. Возникает вопрос: существует ли такой код, и если да, то как его построить?

Коды, не содержащие пробелов

Пусть a, b, c, d — четыре буквы, образующие алфавит рибонуклеиновых кислот. Назовем *диграммой* упорядоченную пару таких букв, например ab, dc, cc. Упорядоченную тройку букв, например cda, bbc, aaa условимся называть *триграммой*. Попробуем начать с триграмм. Предположим, что каждой из 20 букв α, β, γ, ... алфавита протеинов соответствует одна триграмма. Иначе говоря, для того чтобы из 20 аминокислот выбрать одну и при синтезе протеиновой цепочки расположить в нужном месте, необходимо взять три рибонуклеиновые кислоты и выстроить в цепочку в определенной последовательности. Получившийся код будет выглядеть, например, так:

$$abc = \alpha,$$
$$dda = \beta,$$
$$cbd = \gamma,$$
$$bda = \delta$$

и т. д. К сожалению, для дела — перевода с языка рибонуклеиновых кислот на язык протеинов — он непригоден. Такое «рибонуклеиновое» слово, как

$$abcbdabdaddacbdabcbda,$$

по записи в «протеиновом» алфавите выглядело бы следующим образом:

$$\alpha\ \delta\ \delta\ \beta\ \gamma\ \alpha\ \delta$$

Взглянув на 14, 15 и 16-ю буквы первого слова, нетрудно заметить, что из-за слияния триграмм $\gamma = cbd$ и $\alpha = abc$ образовалась триграмма bda, соответствующая в протеиновом алфавите букве δ. Триграмма bda может привести к сбою в синтезе протеина: аминокислота может попасть в цепочку и нарушить стройные пропорции всей сложной постройки молекулы протеина. Так деталь головоломки, схожая по наружным очертаниям с другой деталью, попав не на свое место, может помешать всей сборке. Отсутствие пробелов — требование очень жесткое, и удовлетворить ему нелегко. Возникает математическая задача. Как лучше ее сформулировать?

Из четырех букв a, b, c, d всего можно составить $4^3 = 64$ триграммы. Из них нужно отобрать 20 таких триграмм (образующих систему S), что если триграммы xyz и uvw принадлежат системе S, то ни триграмма yzu, ни триграмма zuv не должны принадлежать системе S. (Переменные x, y, z, u, v, w означают любые из букв a, b, c, d.) Существует ли система S с такими свойствами?

Прежде всего заметим: если триграмма xyz принадлежит системе S, то триграммы yzx и zxy не могут принадлежать S. В частности, триграммы aaa, bbb, ccc, ddd относятся к числу запрещенных. В нашем распоряжении остается $64 - 4 = 60$ триграмм. Из любых трех триграмм xyz, yzx, zxy системе S может принадлежать лишь одна. Следовательно, условие отсутствия пробелов приводит к тому, что система S содержит не более 20 триграмм. Весьма многообещающе выглядит совпадение чисел: мы снова встречаемся с числом 20.

Сформулируем еще раз математическую задачу, решением которой мы сейчас займемся.

Решение задачи

Задача. Назовем триграммой упорядоченную тройку букв, составленную из элементов четырехбуквенного алфавита a, b, c, d. Система S триграмм называется кодом, если из того, что

триграммы xyz и uvw принадлежат S, следует, что

$$\text{триграммы } yzu \text{ и } zuv \text{ не принадлежат } S.$$

Требуется построить все коды, содержащие максимальное число (20) триграмм.

Предположим, что нам удалось построить код S, обладающий всеми нужными свойствами. Рассмотрим следующее.

Утверждение 1. Из каждого набора, состоящего из трех триграмм xyz, yzx, zxy, системе S принадлежит ровно одна триграмма, если только не выполняется равенство $x=y=z$ (в этом случае ни одна из трех триграмм не принадлежит S).

То, что системе S может принадлежать не более одной триграммы из каждого допустимого набора, следует непосредственно из требования об отсутствии пробелов. Поскольку существует лишь двадцать наборов триграмм, то одна триграмма из каждого набора должна принадлежать S, ибо в S должно входить 20 триграмм.

Определение 1. Обозначим S_x множество триграмм, в которых x стоит в середине, L_x — множество букв, стоящих в триграммах из множества S_x слева от x, и R_x — множество букв, стоящих в триграммах из множества S_x справа от x.

Утверждение 2. Если буква u принадлежит множеству L_x, буква v — множеству R_x, причем буквы u, v не равны одновременно x, то триграмма uxv принадлежит системе S.

Действительно, по предположению при заданных q и p существуют определенные триграммы

$$qxv \text{ и } uxp \text{ из } S,$$

причем триграммы

$$xvu \text{ и } vux \text{ не принадлежат } S.$$

Следовательно, по доказанному выше утверждению 1 триграмма

$$uvx \text{ принадлежит } S,$$

что и требовалось доказать.

Определение 2. Введем сокращенный способ записи элементов множества S_x. Выпишем посредине одну под другой четыре раза букву x, слева — одну под другой буквы, принадлежащие множеству L_x, справа — также один под другим элементы множества R_x. Например, если множество L_b состоит из букв a, b, d, а множество R_b — из букв b и c, то S_b будет содержать следующие левые и правые диаграммы:

$$\begin{array}{l} ab \\ bb \\ \ \ bb \\ \ \ bc \\ db \end{array}$$

а наша сокращенная запись будет иметь вид:

$$S_b = \begin{array}{c} abb \\ bbc \\ db \\ b \end{array}$$

Понимать эту запись надлежит следующим образом. Чтобы получить все члены множества S_b, необходимо составить все возможные триграммы, у которых левая буква взята из левого столбца, правая — из правого, а средней буквой выбрана буква b. Запрещенной является лишь триграмма bbb.

Утверждение 3. Если буква x принадлежит либо множеству L_x, либо множеству R_x, то множество S_x пусто (не содержит ни одной триграммы).

Предположим, что буква x не принадлежит ни множеству L_x, ни множеству R_x, а множество S_x содержит по крайней мере одну триграмму, например uxv, причем $u \neq x$. Тогда триграммы uxx и xxu не принадлежат системе S. Следовательно, по утверждению 1 триграмма xux должна принадлежать системе S. Но это противоречит исходному предположению о том, что триграмма uxv принадлежит системе S.

Утверждение 4. Ни одна диграмма pq не может встречаться среди триграмм, образующих систему S, одновременно справа и слева.

Предположим противное. Пусть триграммы

$$pqx \text{ и } ypq \text{ принадлежат } S. \tag{4.0}$$

Если бы в системе S нашлась триграмма, оканчивающаяся буквой y, например uvy, то, сопоставляя ее с триграммой pqx, мы бы пришли к противоречию, поскольку триграмма ypq не могла бы принадлежать S вопреки предположению. Следовательно, в системе S не существует триграммы, оканчивающейся буквой y. Аналогичным образом можно показать, что в системе S не существует триграммы, начинающейся с буквы x. Отсюда следует, что $x \neq y$.

Далее, для каждого z триграммы

$$xxz \text{ и } xzx \text{ не принадлежат } S.$$

По утверждению 1 при $z \neq x$ отсюда следует, что триграмма

$$zxx \text{ принадлежит } S. \tag{4.1}$$

Кроме того, триграммы

$$zyy \text{ и } yzy \text{ не принадлежат } S.$$

По утверждению 1 при $z \neq y$ это означает, что триграмма

$$yyz \text{ принадлежит } S. \tag{4.2}$$

В частности, подставляя в (4.1) букву y, а в (4.2) букву x вместо z, получаем: триграмма

$$yxx \text{ принадлежит } S, \tag{4.3}$$

$$yyx \text{ принадлежит } S. \tag{4.4}$$

Из (4.1) и (4.4) следует, что при $z \neq x$ триграмма

$$yxz \text{ не принадлежит } S, \tag{4.5}$$

а из (4.3) и (4.2) — что при $z \neq y$ триграмма

$$zyx \text{ не принадлежит } S. \tag{4.6}$$

Пользуясь утверждением 1, заключаем из (4.5) и (4.6), что при любых $z \neq x$, $z \neq y$ триграмма

$$xzy \text{ принадлежит } S.$$

Но это противоречит доказанному ранее утверждению о том, что ни одна триграмма из S не может начинаться с буквы x и оканчиваться буквой y.

Следовательно, высказанное нами предположение (4.0) неверно. Тем самым утверждение 4 доказано.

Утверждение 5. Буква x не может одновременно принадлежать множествам L_x и R_x.

Действительно, в противном случае диграмма xx встречалась бы среди триграмм, образующих систему S одновременно и справа, и слева (вопреки утверждению 4).

Следовательно, множество S_b, приведенное нами в качестве примера в определении 2, не существует.

Утверждение 6. Если x принадлежит R_x и u принадлежит R_x (причем $u \neq x$), то u принадлежит также L_x.

По предположению xx и xy встречаются в системе S как правые диаграммы. Следовательно, по утверждению 4 комбинации xx и xy не могут входить в триграммы из S как левые диаграммы, и, таким образом, триграммы

$$xxu \text{ и } xux \text{ не принадлежат } S.$$

Но тогда по утверждению 1 триграмма

$$uxx \text{ принадлежит } S.$$

Но это и означает, что буква u принадлежит L_x.

Утверждение 7. Пусть S_x — не пустое множество. Тогда по утверждению 3 буква x должна принадлежать либо множеству R_x, либо множеству L_x. Предположим для определенности, что x принадлежит множеству R_x. Обозначим s_x, l_x, r_x число элементов в множествах S_x, L_x, R_x. Из утверждения 6 следует, что

$$r_x \leqslant l_x + 1.$$

По утверждению 2

$$s_x = l_x \cdot r_x.$$

Обозначим s'_x число триграмм из S_x с двумя одинаковыми буквами. Тогда

$$s'_x = l_x + r_x - 1.$$

Если x принадлежит R_x, то r_x, l_x, s_x и s'_x могут принимать самое большее следующие значения:

$r_x =$	4	3	2	1	3	2	1	2	1	0
$l_x =$	3	3	3	3	2	2	2	1	1	0
$s_x =$	12	9	6	3	6	4	2	2	1	0
$s'_x =$	6	5	4	3	4	3	2	2	1	0

(Нули в последнем столбце соответствуют пустому множеству S_x.)

Система S содержит ровно 20 триграмм, 12 из них содержат по 2 одинаковых буквы. Таким образом, должны выполняться соотношения

$$s_a + s_b + s_c + s_d = 20,$$
$$s'_a + s'_b + s'_c + s'_d = 12.$$

С учетом их мы получаем для четырех чисел s_d, s_c, s_b, s_a лишь следующие наборы допустимых значений:

I;	12,	6,	2,	0;
II:	12,	6,	1,	1;
III:	12,	4,	4,	0;
IV:	9,	9,	2,	0;
V:	9,	9,	1,	1

и, разумеется, все наборы, возникающие из наборов I—V при перестановке букв a, b, c, d.

Рассмотрим случай I: $s_d = 12$, $s_c = 6$. При $s_d = 12$ находим: $r_d = 4$, $l_d = 3$ и

$$S_d = \begin{matrix} ada \\ bdb \\ cdc \\ dd \end{matrix}$$

плюс все возможные перестановки левых и правых букв. В триграммы рассматриваемого множества S_d диграммы ad, bd, cd входят слева. По утверждению 4 это означает, что они не могут входить ни в одну диграмму системы S справа. Следовательно, буква d может входить справа в триграммы, образующие систему S, лишь в том случае, если эти триграммы принадлежат множеству S_d. Кроме того, da, db, dc входят как правые диграммы в триграммы множества S_d. Следовательно, da, db, dc не могут входить как левые диграммы в триграммы системы S. Таким образом, буква d не может входить в триграммы системы S слева и вообще не входит в триграммы системы S, не принадлежащие множеству S_d.

При $s_c = 6$ находим: $r_d = 3$, $l_d = 2$ и

$$\begin{array}{c} aca \\ bcb \\ cc \\ c \end{array}$$

плюс все возможные перестановки левых и правых букв. По аналогии с приведенными выше рассуждениями заключаем, что буква c не входит в триграммы, образующие множества S_b и S_a. В случае I при $s_b = 2$, $r_d = 2$ и $l_d = 1$ множество S_b может иметь лишь следующий вид:

$$\begin{array}{c} aba \\ bb \\ b \\ b \end{array}$$

плюс все возможные перестановки левых и правых букв, а множество S_a пусто. В случае II, когда $s_b = s_a = 1$, множества S_b и S_a имеют следующий вид:

$$\begin{array}{cc} ab & aa \\ bb & ba \\ b & a \\ b & a \end{array}$$

Правые и левые буквы здесь также разрешается переставлять, но лишь одновременно в S_a и в S_b, поскольку в противном случае диграмма ab входила бы в триграммы системы S справа и слева.

Рассмотрим случай III с $s_c = 4$. С точностью до перестановки правых и левых букв триграмм можно предположить, что множество R_c содержит букву c. Если буква b принадлежит R_c, то по утверждению 6 буква b одновременно принадлежит и множеству L_c. Следовательно, триграммы, образующие систему S, не могут содержать диграмму cb слева и диграмму bc справа. Это означает, что в триграммах системы S буква c не может стоять рядом с буквой b. Таким образом, соседями буквы b могут быть лишь буквы a и b. Но тогда наибольшее значение s_b не превышало бы 2, в то время как в случае III должно выполняться равенство $s_b = 4$. Из полученного противоречия следует, что буква b не принадлежит множеству R_c, и множества S_c и S_b состоят из элементов

$$\begin{array}{rr} aca & aba \\ bc\ \ & bb \\ cc & cb\ \ \\ c & b \end{array}$$

плюс все возможные перестановки правых и левых букв.

В случае IV выполняются равенства $s_d = s_c = 9$.

Предположим, что множество R_d содержит букву d. Если бы буква c принадлежала множеству R_d, то по утверждению 6 она принадлежала бы и множеству L_d. Но тогда, как сразу видно, буква d входила бы в триграммы множества S_c и справа, и слева. Поскольку буква c входит либо в множество L_c, либо в множество R_c, то $s_c \leqslant 6$. Это неравенство противоречит равенству $s_c = 9$, которое должно выполняться в рассматриваемом случае, поэтому буква c не может принадлежать множеству R_d. Итак, множества S_d и S_c содержат следующие элементы:

$$\begin{array}{rr} ada & aca \\ bdb & bcb \\ cd\ \ & cc \\ dd & dc\ \ \end{array}$$

плюс все возможные перестановки правых и левых букв. В триграммы системы S диграммы ad, bd, ac, bc не могут входить справа, а диграммы da, db, ca, cb — слева. Следовательно, буквы c и d не могут входить в триграммы системы S, не принадлежащие множествам S_c и S_d. Таким образом, при $s_b = 2$, $s_a = 0$ множество S_b содержит элементы

$$\begin{array}{r} aba \\ bb \\ b \\ b \end{array}$$

а в случае V при $s_b = s_a = 1$ множества S_b и S_a — элементы

$$\begin{array}{cc}
ab & aa \\
bb & ba \\
b & a \\
b & a
\end{array}$$

плюс обычные перестановки правых и левых букв.

Следствие. Итак, соберем воедино все полученные результаты. Полное решение задачи имеет следующий вид:

$$\begin{array}{ccccc}
\mathrm{I} & ada & aca & aba & a \\
 & bdb & bcb & bb & a \\
 & cdc & cc & b & a \\
 & dd & c & b & a \\
\mathrm{II} & ada & aca & ab & aa \\
 & bdb & bcb & bb & ba \\
 & cdc & cc & b & a \\
 & dd & c & b & a \\
\mathrm{III} & ada & aca & aba & a \\
 & bdb & bc & bb & a \\
 & cdc & cc & cb & a \\
 & dd & c & b & a \\
\mathrm{IV} & ada & aca & aba & a \\
 & bdb & bcb & bb & a \\
 & cd & cc & b & a \\
 & dd & dc & b & a \\
\mathrm{V} & ada & aca & ab & aa \\
 & bdb & bcb & bb & ba \\
 & cd & cc & b & a \\
 & dd & dc & b & a
\end{array}$$

Для получения полного списка допустимых триграмм необходимо произвести еще дополнительные операции:

1) произвольные перестановки букв a, b, c, d;

2) перестановки правых и левых букв в каждом множестве S_x, причем в случае II переставлять буквы в S_b и в S_a разрешается лишь одновременно, в случае III такое же ограничение накладывается на перестановку левых и правых букв во множествах S_c и S_b,

в случае IV — во множествах S_d и S_c и в случае V — во множествах S_c и S_a.

Таким образом, в случаях I и II триграммы распадаются на 8 типов, а в случаях III, IV, V — на 4 типа. С точностью до перестановок букв a, b, c, d всего получается 28 решений.

Утверждение 8. Мы доказали лишь, что решения задачи (коды, содержащие максимальное число триграмм), если они существуют, содержатся среди построенных нами 28 систем триграмм. Необходимо еще убедиться в том, что все 28 систем действительно являются кодами без пробелов. Проверка этого условия настолько легка, что мы предоставляем ее читателю.

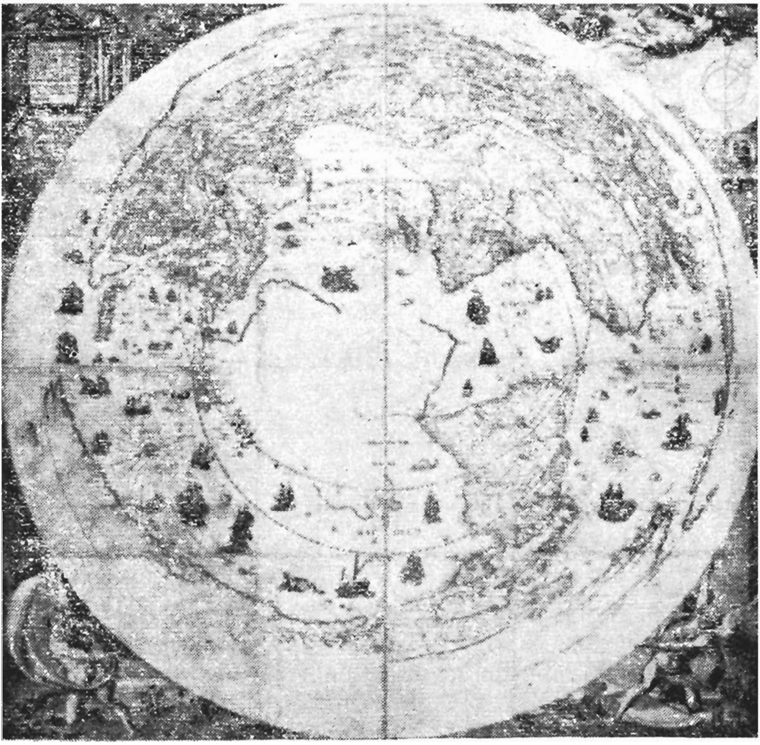
Заключение

Можно ли утверждать, что одно из 28 полученных нами решений служит азбукой жизни? Этот вопрос относится к компетенции не математика, а самой природы. Поставив эксперимент, мы можем задать ей вопросы, и если они разумны, то природа даст на них разумные ответы.

Можно лишь с уверенностью сказать, что в действительности передача наследственной информации происходит не столь просто. Полученные нами результаты просты лишь потому, что мы исходим из простых предположений. Но так ли просто действует природа? Нет, уже сейчас удалось выяснить, что наследственные признаки передаются более сложным путем. Найденное нами решение слишком красиво, чтобы быть верным. Но об этом вряд ли следует сожалеть: ведь мы занялись решением математической задачи лишь для того, чтобы поразвлечься, а умение развлекаться, несомненно, входит в азбуку жизни.

Глава 5

Искусство рисовать плохо



Непрерывная деформация букв

С тех пор как мы научились читать, кому из нас не случалось принимать букву X за букву U. Но в этой главе нам придется позабыть почти все, чему нас учили в первом классе. Например, стоит лишь распрямить букву J, как она превратится в букву I. Потрудившись чуть больше, мы сумеем превратить в I буквы C, S, и U. Стоить лишь отбросить поперечные черточки на концах букв L и V, как их также можно будет превратить в I. Перегнув букву I дважды, мы получим буквы Z или N, а перегнув трижды — буквы M или W.

Дело в том, что буквы

C, I, J, L, M, N, S, U, V, W, Z (1)

обладают одним важным свойством: они *топологически эквивалентны*. Их можно наложить одну на другую. Правда, для этого окажется недостаточным ни параллельный сдвиг, совмещающий конгруэнтные треугольники, ни преобразование подобия, растягивающее или сжимающее геометрические фигуры, ни даже проектирование, переводящее окружность в эллипс. Наложить любые две буквы (1) друг на друга нам удастся лишь в том случае, если мы их непрерывно деформируем.

С буквой O у нас ничего бы не вышло. Если бы мы захотели наложить ее на букву I, то нам бы пришлось сначала либо перерезать букву O в какой-то точке, либо сплющить ее так, чтобы различные точки буквы O совпали с одной и той же точкой буквы I. Но ни разрезать, ни сплющивать буквы не разрешается. Буква O топологически не эквивалентна букве I, а также всем остальным буквам ряда (1). Наоборот, буквы

D, O (2)

топологически эквивалентны. К другому топологическому типу принадлежат буквы

E, F, G, T, Y. (3)

Они также топологически эквивалентны друг другу. Например, достаточно взять букву E за «перекладинку» и повернуть всю букву вправо на 90°, как, растянув или сжав отдельные ее элементы, мы получим букву T. Еще проще превратить букву E в букву F. Взяв букву G за горизонтальную черточку и распрямив искривленную часть буквы так, чтобы она расположилась по вертикали вниз, мы снова получим букву T. Распрямив «рога», торчащие у буквы Y, мы также без труда преобразуем ее в букву T. И при всех этих превращениях мы ничего не склеивали и не разрезали. Если у исходной буквы какие-то две точки не совпадали, то и у конечной буквы соответствующие им две точки не сливались в одну. При непрерывной деформации одной буквы в другую разрывы и склеивания находятся под строгим запретом. Именно поэтому ни одну из букв (3) нельзя превратить ни в одну из букв (1) или (2). Буква T и другие топологически эквивалентные ей буквы обладают одной отличительной особенностью: точкой, из которой выходят три отрезка. Ни у одной из букв (1) или (3) такой точки нет. Но это еще не все. У букв

$$\text{K, X} \tag{4}$$

имеется точка, из которой выходят четыре отрезка. Поэтому буквы K, X топологически отличны от букв (1), (2), (3), хотя они топологически эквивалентны друг другу. У буквы

$$\text{H} \tag{5}$$

есть 2 точки, из которых выходят по три отрезка. Эта буква образует отдельный топологический тип: она топологически эквивалентна лишь самой себе. Буква

$$\text{P,} \tag{6}$$

так же как и буквы (2), содержит замкнутый участок. Она топологически эквивалентна окружности на ножке и поэтому отлична от букв (2). Буквы

$$\text{Q, R} \tag{7}$$

можно представить в виде окружности с двумя ножками, выходящими из одной и той же точки, букву

$$\text{A} \tag{8}$$

— в виде окружности с двумя ножками, выходящими из двух различных точек. Наконец, буква

$$\mathrm{B} \tag{9}$$

топологически эквивалентна двум склеенным в одной точке окружностям.

Геометрия и топология

Геометрия означает измерение Земли, и, хотя люди довольно быстро научились абстрагироваться от задач, с которыми они сталкивались на измеряемой Земле, прошло две с половиной тысячи лет, прежде чем геометры вновь занялись измерениями. Но тут как раз

Рис. 31.

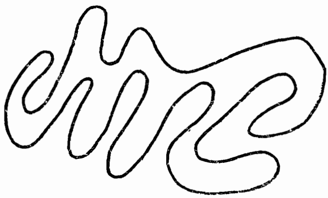


Рис. 32.

появились геометрии, в которых еще шла речь о точках, прямых и плоскостях, но ни словом не упоминались расстояния и углы, подлежащие измерению. В этих геометриях по-прежнему пользовались линейкой, но без сантиметровой шкалы. В нашем веке геометры, уже успевшие окончательно отказаться от линейки, сначала робко, а затем во всю мощь своего голоса заявили о себе как о ... плохих художниках, не умеющих отличать отрезок кривой (рис. 31) от прямолинейного отрезка, а замкнутую линию, изображенную на рис. 32, — от окружности. Новую разновидность геометрии назвали *топологией.*

Карты

Игра, которой мы забавлялись в начале этой главы, переводя путем непрерывной деформации одну латинскую букву в другую, называется топологией кривых. Поднимемся теперь на одну ступень и займемся топологией поверхностей.

На поверхность сферы мы смотрим как на прототип всех поверхностей, но наш вывод слишком поспешен. Долгое время люди считали, что живут на огромном круглом диске. И действительно, окружность — прекраснейшая из фигур, которые можно построить на плоскости, если кто-то захочет начертить карту мира. Гораздо сложнее начертить карту замкнутой поверхности Земли. О том, как это делается, мы немного рассказали в главе 1. Землю мысленно разрезают вдоль экватора на северное и южное полушария или вдоль

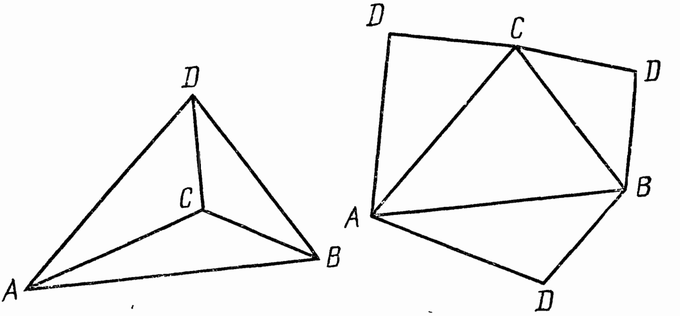


Рис. 33.

нулевого меридиана — на западное и восточное полушария, изображают оба полушария на плоскости в виде кругов и располагают обе картинки рядом. Точки, лежащие на экваторе или на нулевом меридиане, оказываются при этом дважды нанесенными на карту: им соответствуют граничные окружности полушарий. Условные обозначения («оцифровка») на обрезе карты показывают, какие точки следует отождествлять. Карта двух полушарий представляет собой не что иное, как модель поверхности сферы. На картах, вычерченных в меркаторской проекции, поверхность Земли изображается в виде прямоугольника. Все верхнее основание прямоугольника соответствует одной-единственной точке земной поверхности — Северному полюсу, все нижнее основание — Южному полюсу. Точки на правом и левом обрезе карты совпадают, если лежат на одной и той же географической широте. Если нам потребуется изобразить на плоскости модель тетраэдра, поверхность которого также замкнута, то проще всего поставить тетраэдр одной гранью на плоскость и развернуть на нее три остальные грани (рис. 33). Карта (или развертка) поверхности тетраэдра имеет вид шестиугольника, и отождествить участки границы не представляет труда. Наконец, можно не ограничиваться составлением одной обзорной карты поверхности и составить атлас с подробными картами окрестностей ее точек. Частичное или полное соответствие между отдельными картами атласа и местностью помогает устанавливать различные условные обозначения, названия населенных пунктов, рек, морей, а также координатную сетку — меридианы и параллели.

Примеры поверхностей

Убедившись в том, что земная поверхность по форме отличается от круглого диска, люди обратили свои помыслы к прекраснейшей из замкнутых поверхностей — к сфере. С точки зрения топологии большинство тел, с которыми нам приходится сталкиваться, имеют форму сферы (причем не только геометрические тела, но и наше собственное тело). Даже говоря о микромире, человек представляет себе электроны, протоны и кванты света в виде крохотных шариков.

Разумеется, среди наших соседей по Вселенной не обошлось без возмутителя спокойствия. Мы имеем в виду кольца Сатурна. Предположим, что обитатель Сатурна захотел начертить карту своего мира. Кольцо, на котором он обретается, обитатель Сатурна разрезал бы вдоль меридиана. Но этого недостаточно: получился бы цилиндр, и для того, чтобы его поверхность можно было бы развернуть на плоскости, необходимо провести еще один разрез вдоль образующей, то есть вдоль параллели кольца, на котором обитает наш друг с Сатурна (рис. 34). Карта мира получилась бы в форме прямоугольника со сторонами, отождествленными самым неожиданным образом (рис. 35). Точки на правом и левом обрезах карты совпадают, если лежат на одной и той же сатурнографической широте; точки на верхнем и нижнем обрезах совпадают, если лежат на одной и той же сатурнографической долготе.

Математики называют такую поверхность *тором.* Устроен тор иначе, чем сфера. Например, на торе можно начертить океан, а в океане расположить остров с двумя замкнутыми береговыми линиями, не имеющими общих точек. Такой остров изображен на рис. 36:

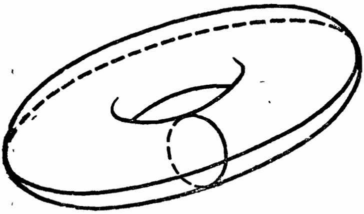


Рис. 34.

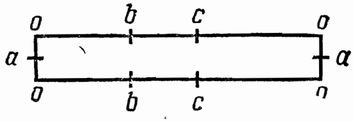


Рис. 35.

суша заштрихована, а океан оставлен белым (следует иметь в виду, что океан состоит лишь из одного «куска» — он занимает *связную* часть поверхности тора). На сфере что-либо подобное невозможно. Остров в

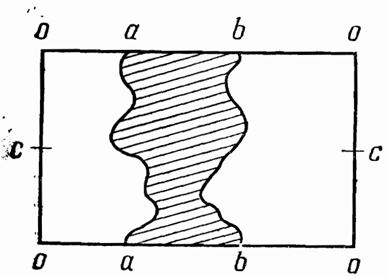


Рис. 36.

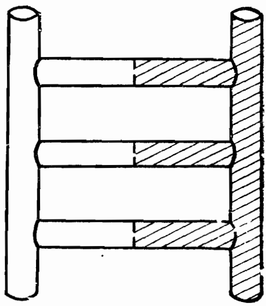


Рис. 37.

океане с озером, расположенным в центре, имеет 2 берега: внутренний и внешний. Часть земной поверхности, занимаемая в этом случае водой, несвязна.

Еще более сложный вид имеет поверхность приставной лестницы с $p+1$ ступеньками при $p>1$. (При $p=0$ и $p=1$ мы получаем лестницы с 1 и 2 ступеньками, поверхность которых топологически эквивалентна поверхности сферы и тора. «Рога», которыми оснащены деформированная сфера и тор, с топологической точки зрения несущественны.) Это —

нечто новое. На такой поверхности можно начертить остров в океане с $p+1$ береговыми линиями, не имеющими общих точек. Такой остров изображен на рис. 37, где справа показана суша (она заштрихована), а слева — океан, и ясно видны береговые линии, отделяющие воду от суши.

В топологии такие поверхности принято называть *поверхностями рода* p. Сфера — поверхность рода 0, тор — поверхность рода 1.

Поверхности рода p можно представить себе следующим образом. Просверлим в шаре p сквозных непересекающихся каналов. Поверхность оставшейся части шара, как нетрудно понять, будет поверхностью рода p. Следовательно, поверхность человеческого тела, если иметь в виду не только поверхность кожного покрова, но и слизистую оболочку пищевода, желудочно-кишечного тракта и другие каналы, может служить примером поверхности весьма высокого рода. Можно указать и многие другие реально существующие тела, ограниченные отнюдь не сферообразной поверхностью.

Но продолжим наше знакомство с замкнутыми поверхностями иного (ненулевого) рода, чем сфера. Одной из таких поверхностей является так называемая *проективная плоскость*. На обычной плоскости одни пары прямых пересекаются, другие — параллельные — не имеют точек пересечения. Трудами многих исследователей удалось достичь более приятного однообразия. Пополним плоскость несобственными, или бесконечно удаленными, точками. На каждой прямой лежит *одна* несобственная точка. Параллельные прямые не имеют других общих точек, кроме несобственных, причем всем параллельным прямым принадлежит одна и та же несобственная точка. Так, все горизонтальные прямые на рис. 38 «проходят» через одну и ту же несобственную точку, все наклонные прямые также проходят через одну несобственную точку, отличную от той, которая принадлежит горизонтальным прямым. Вертикальные прямые проходят через третью несобственную точку, не совпадающую ни с одной из двух первых несобственных точек, и так далее. Все несобственные точки лежат на одной — несобственной — прямой.

На плоскости, пополненной несобственной прямой, любые две несовпадающие прямые пересекаются ровно в одной точке: непараллельные прямые — в собственной, или конечной, точке плоскости, параллельные — в несобственной точке. Несобственная прямая

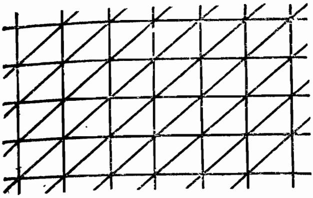


Рис. 38.

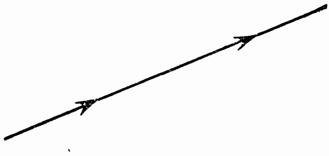


Рис. 39.

пересекается с любой собственной прямой в принадлежащей той прямой несобственной точке. Прямые на проективной плоскости замкнуты. Двигаясь по прямой на проективной плоскости в бесконечность, мы возвратились бы в исходную точку. Наглядно можно представить себе, что мы пришли бы туда, откуда вышли, с другой стороны, из противоположной бесконечности (рис. 39).

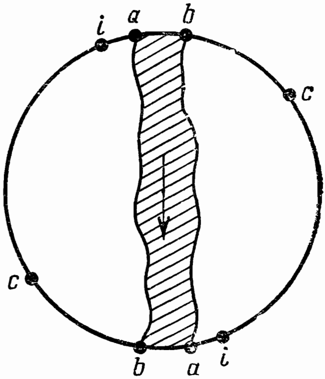


Рис. 40.

Такова проективная плоскость. Чтобы начертить ее карту, подвергнем плоскость небольшой деформации. Стянув плоскость, переведем несобственную прямую в пределы конечного, где она как бы станет своеобразным горизонтом. Диаметрально противоположные точки новоявленного горизонта отождествлены. Двигаясь по прямой в одну сторону и «скрывшись за горизонтом», мы тотчас же «появимся» из-за горизонта с противоположной стороны. Карта проективной плоскости выглядит так, как показано на рис. 40, — круг с отождествленными диаметрально противоположными точками граничной окружности. Нельзя не признать, что точки границы склеены между собой (отождествлены) весьма хитроумным способом. Однако реализовать модель проективной плоскости обычными средствами невозможно, абстрактная же ее модель безупречна.

Проективная плоскость совсем не похожа на те поверхности, с которыми нам приходилось встречаться раньше. Она замкнута, но замкнута иначе, чем привычные нам поверхности. Нарисуем на карте проективного мира узкую речку (на рис. 40 она заштрихована). У проективной реки есть лишь один берег! Путешественник, совершающий кругосветное путешествие на теплоходе, отправляясь из пункта a (в верхней части рис. 40), видел блеск маяка i по правому борту. Вернувшись назад (добравшись до пункта a в нижней части рис. 40), он увидит блеск того же маяка i по левому борту. Встречающие его друзья, не отлучавшиеся из дому, к своему удивлению заметят, что на их приятеле пагубно сказались тяготы дальнего рейса. Он стал буквально «сам не свой»: из обычного человека, предпочитавшего действовать правой рукой, превратился в левшу, сердце у него бьется справа, стрелки часов движутся в противоположную сторону, а чтобы отдраить иллюминаторы, барашки болтов ему придется крутить в ту же сторону, в которую раньше их крутили при задраивании. Чтобы возвратиться в старое состояние, путешественнику пришлось бы проделать еще одно кругосветное путешествие по проективной плоскости.

Поверхности, на которых происходят столь необычные явления, называются *неориентируемыми*. Проективная плоскость — не единственная неориентируемая поверхность. Каждой замкнутой ориентируемой поверхности (а такие поверхности мы полностью перечислили) соответствует неориентируемый «довесок». Мы ограничимся упоминанием об этом небезынтересном обстоятельстве и перейдем сейчас к другому вопросу.

Поверхности в целом

А что, собственно, мы называем поверхностью?

На этот вопрос мы успели ответить еще до того, как он был задан: отличительная особенность поверхности состоит в том, что окрестность любой ее точки можно отобразить на карту.

Смысл данного нами определения поверхности полезно уяснить на примерах «неповерхностей». Сфера с торчащими из нее «хвостиками» — кривыми — уже не будет поверхностью: для того чтобы изобразить на карте окрестность точки, в которой к сфере прикрепляется хвост, нам пришлось бы подвесить аналогичный «придаток» и к соответствующей точке карты, а карту с хвостом даже при желании нельзя назвать картой. Две сферы, склеенные в одной точке, также не образуют поверхности, поскольку в противном случае нам пришлось бы отождествить на двух обычных картах по одной точке, окрестности которых не соответствовали бы друг другу. (Если кто-нибудь читает за завтраком газету и случайно уронит капельку меда между двумя страницами, то газетные листы склеются, но и такую «самоделку» из бумаги нельзя назвать картой.) Ясно также, что если мы хотим получить поверхность, то две сферы не рекомендуется склеивать вдоль какой-нибудь линии или участка поверхности, поскольку окрестность линии склейки или края поверхности «контакта» двух сфер нельзя нанести на карту.

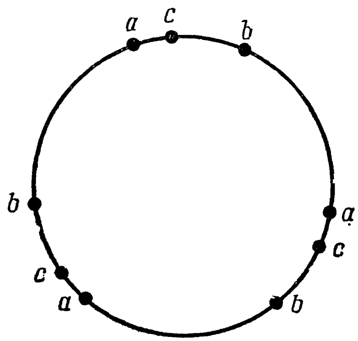


Рис. 41.

Аналогично обстоит дело и со следующим примером (рис. 41). Отождествим те точки, расположенные на границе круглого диска, угловые координаты которых отличаются на 120° или на 240°. Точки на краю диска при этом окажутся как бы склеенными в группы, каждая из которых содержит по 3 точки. То, во что превращается диск при таком отождествлении граничных точек, не будет поверхностью: кривая на «поверхности» имеет 3 берега вместо двух, как полагается на карте.

Итак, поверхность — это атлас карт с определенными правилами отождествления некоторых точек.

В окрестности любой своей точки поверхность должна быть устроена так, чтобы эту окрестность можно было нанести на карту. Занимаясь составлением карт, мы встретим различные типы точек, которые небесполезно было бы научиться распознавать. *Внутренней* называется точка множества, сколь угодно малая окрестность которой целиком содержится во множестве. Точка множества называется *граничной*, если ее окрестность содержит как точки самого множества, так и точки, не принадлежащие множеству (если множество имеет границу, то граничная точка принадлежит ей). Наконец, точка называется *изолированной*, если ее окрестность целиком состоит из точек, не принадлежащих множеству. Разумно потребовать, чтобы каждая точка множества по крайней мере на одной карте была внутренней точкой. Это позволило бы единым взглядом охватывать ее окрестность. Если точка встречается на нескольких картах, то ее окрестности в том виде, как они приведены на картах, отождествляются так, как этого требуют правила отождествления. (Если точка оказывается внутренней на одной карте и граничной на другой, то отождествить удается лишь часть окрестности внутренней точки и окрестность граничной точки.) Разумеется, внутренние точки одной и той же карты никогда не отождествляются, поскольку при таком отождествлении карта перестала бы быть картой.

Наконец, необходимо еще одно требование. Поверхность должна быть связной, а не состоять, например, из двух отдельных сфер. Как это проявляется в строении атласа? Начнем с карты 1. На ней нам может встретиться точка, которая нанесена также на карты 3, 7, 11. На карте 3 мы обнаружим точку, представленную на картах 1, 11, 12, на карте 7 — точку, встречающуюся на картах 1, 12, 14, на карте 11 — точку, нанесенную на картах 1, 3, 14, 19. Короче говоря, к карте 1 примыкают (перекрываясь с ней) карты 3, 7, 11, к карте 3 — карты 1, 11, 12, к карте 7 — карты 1, 12, 14 и так далее. Продолжая путешествие по страницам атласа достаточно долго, мы сможем в конце концов добраться до любой карты атласа. В этом и состоит требование, которое нам предстоит наложить.

Атлас может содержать бесконечно много карт. Может оказаться, например, при рассмотрении поверхности приставной лестницы с бесконечно большим числом перекладин, что обойтись конечным числом карт просто нельзя. Если же поверхность можно отобразить в атласе, состоящем из конечного числа карт, то такая поверхность называется *замкнутой.*

Многообразия

В том, что мы поведали читателю, по существу содержатся точные определения топологии поверхностей и, в частности, замкнутой поверхности. При желании и тому, и другому определению можно было бы придать сжатый вид, но вместо этого мы хотя бы ненадолго должны заглянуть в высшие размерности.

В главе 1 мы познакомились с замкнутым пространством: модель его нам удалось построить, взяв два шара и отождествив их поверхности. Добравшись во время путешествия по такому пространству до границы одного шара, мы тем самым оказались бы на границе другого: в само́м замкнутом пространстве никаких граничных поверхностей не существует.

Если мы захотим составить описание пространства теми же средствами, что и описание поверхностей, то плоские карты нам придется заменить объемными, например глобусами или кубами, и составить атлас таких трехмерных карт. В остальном атлас пространства должен был бы удовлетворять тем же требованиям, которые мы установили для атласов поверхностей. То, что у нас получилось бы, в топологии принято называть *многообразием.* Примером многообразия может служить замкнутое пространство из главы 1 (правда, атлас, состоящий лишь из двух сфер, необходимо несколько расширить, поскольку каждая точка пространства должна изображаться внутренней точкой по крайней мере на одной «карте»).

Рассмотрим другой пример многообразия. Возьмем шар, вырежем из него концентрический шар меньшего радиуса и отождествим точки внутренней и внешней поверхностей образовавшейся полой сферической оболочки. Например, можно отождествить точки внутренней и внешней поверхностей, лежащие на одном диаметре по одну сторону от центра (таковы точки a и a' на рис. 42), или точки внутренней и внешней поверхностей, лежащие на одном диаметре по разные стороны от центра (таковы точки a и a'' на рис. 42). Нетрудно показать, что эти два способа отождествления

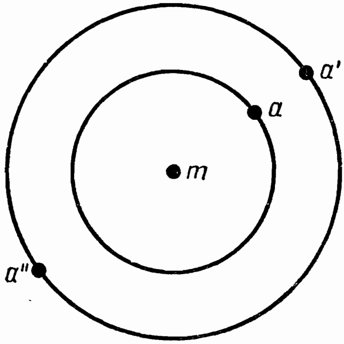


Рис. 42.

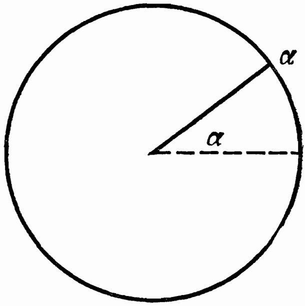


Рис. 43.

точек на граничных поверхностях сферической оболочки приводят к различным многообразиям.

Отображения окружности на себя

Поясним теперь, что такое топология, на другом примере.

Рассмотрим окружность C и отобразим ее на самое себя, то есть сопоставим каждой точке c окружности C другую или ту же самую точку c' окружности C. Точка c называется прообразом точки c' (относительно заданного нами отображения), а точка c' — образом точки c (относительно того же отображения). Образ окружности C совпадает с самой окружностью C, хотя образы отдельных точек не обязательно должны совпадать с прообразами: при отображении точки могут перемещаться по окружности. Каждой точке c окружности C можно поставить в соответствие угол α, который образован с некоторым выбранным направлением радиусом, проведенным в точку c из центра окружности. Соответствие между точками окружности и углами можно обратить, и каждому углу α

сопоставить определенную точку c окружности C. Такое соответствие называется *взаимно однозначным.*

Отображение, при котором каждый угол α переходит сам в себя, называется тождественным. Менее просто устроено отображение, переводящее при любом α угол α в угол $\alpha + 10^\circ$. Оно представляет собой не что иное, как поворот окружности на 10° вокруг центра. Совсем иначе выглядит отображение, переводящее любой угол α в 2α, то есть 0° в 0°, 30° в 60°, 180° в 360°, но 210° в $420^\circ = 60^\circ$ и так далее. В отличие от прыдыдущих отображений удвоение угла не взаимно однозначно: различные точки-прообразы переходят в одну и ту же точку — образ. Отображение, переводящее любой угол α в 3α, также не взаимно однозначно: прообразы, углы которых отличаются на 120° или на 240°, при утроении угла переходят в одну и ту же точку.

Отображение, переводящее угол α в угол $n\alpha$, можно рассматривать при любом целом, но непременно целом, значении n. Например, значение $n = {}^1/_2$ недопустимо потому, что углы 0° и 360° соответствуют одной и той же точке окружности и должны иметь один и тот же образ, а при «отображении» $\alpha \to {}^1/_2\alpha$ они переходили бы в две различные точки окружности, соответствующие углам ${}^1/_2 \cdot 0^\circ = 0^\circ$ и ${}^1/_2 \cdot 360^\circ = 180^\circ$. Зато числу n разрешается придавать любые целые отрицательные значения. При $n = -1$ мы получаем отображение, переводящее любой угол α в угол $-\alpha$. Это не что иное, как отражение окружности C относительно ее диаметра. Такое отображение взаимно однозначно. При $n = -2$ мы снова получаем не взаимно однозначное отображение, но отличное от отображения при $n = 2$. Аналогичное утверждение справедливо при любом целом значении n. Если точка, соответствующая углу α, пробегает окружность C один раз, то ее образ, соответствующий углу $n\alpha$, пробегает окружность n раз. Если представим себе окружность C в виде резинового колечка, надетого на круглую деревяшку, то после отображения $\alpha \to n\alpha$ кольцо окажется натянутым на деревяшку в n слоев, навитых либо в одном, либо в другом направлении. Значение $n = 0$ также допустимо. При $n = 0$ вся окружность C отображается в одну точку с $\alpha = 0^\circ$.

Разумеется, отображениями $\alpha \to n\alpha$ отнюдь не исчерпываются все отображения окружности C на себя. Натяжение резинового кольца, облегающего круглую деревяшку, может меняться от одной точки к другой. Колечко может также собираться в складки. Нельзя лишь разрывать его. На математическом языке это требование выражают так: отображение должно быть *непрерывным*, близкие прообразы при отображении должны переходить в близкие образы.

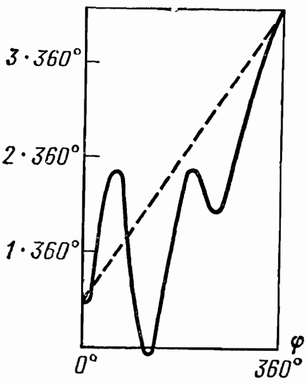


Рис. 44.

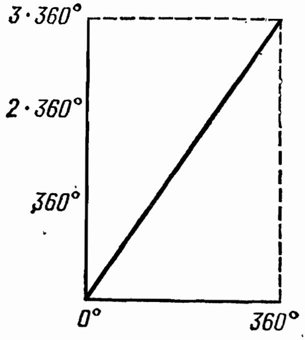


Рис. 45.

Обозначив непрерывное отображение f, изобразим его в виде графика (рис. 44). По горизонтальной оси отложим углы от 0 до 360°. При рассмотрении образов удобно выходить за эти пределы и допускать углы не только больше 360°, но и меньше 0°. Если $f(x)$ — образ угла α, то отложим по вертикали отрезок длиной $f(\alpha)$ (вверх при $f(\alpha) > 0$ и вниз при $f(\alpha) < 0$) и отметим его конец. Начав с угла 0°, построим такие отрезки для всех углов. Концы отрезков образуют некоторую кривую. Если $f(\alpha) > 360°$, то мы не будем приводить угол к интервалу, заключенному между 0 и 360°, а продолжим отрезок за 360°, удлинив его в нужное число раз. Линию $f(\alpha) = 0°$ также разрешается пересекать, переходя при этом из области положительных $f(\alpha)$ в область отрицательных $f(\alpha)$.

График нашего отображения будет непрерывной кривой. Для того чтобы $f(0°)$ и $f(360°)$ соответствовали одной и той же точке окружности C, функция f должна иметь вид

$$f(\alpha) = n\alpha \quad (n \text{ — целое число}).$$

При таком выборе f угол $n \cdot 360°$ всегда кратен углу $360° = 0°$. График отображения f: $\alpha \to n\alpha$ имеет вид прямой (график на рис. 45 соответствует $n = 3$). При отрицательном значении n прямая не поднималась бы, а падала вниз; при $n = 0$ прямая была бы горизонтальной. Отображения $f(\alpha) = n\alpha$ в отличие от отображения, график которого представлен на рис. 44, не имеют складок. Нетрудно видеть, что колебаниям кривой на рис. 44 соответствуют складки отображения f.

Степень отображения

Нам хотелось бы каким-нибудь образом построить классификацию отображений окружности C на себя. Два отображения принадлежат к одному и тому же классу, если их можно непрерывно перевести друг в друга, то есть если одно отображение получается из другого при непрерывном преобразовании. Этот принцип классификации отображений, предложенный голландским математиком Л. Э. Я. Брауэром (1910), оказался необычайно плодотворным в топологии. Именно с него по существу начался современный этап в развитии этой области науки.

Непрерывному преобразованию отображения соответствует непрерывная деформация его графика. Подвергая график отображения непрерывной деформации, необходимо лишь следить за тем, чтобы углы $0°$ и $360°$ имели один и тот же образ. Следовательно, $f(0°)$ и $f(360°)$ надлежит сдвигать одновременно и на одну и ту же величину.

Ясно, что значения углов $f(0°)$ и $f(360°)$ связаны между собой соотношением

$$f(360°) = f(0°) + n \cdot 360°.$$

Целое число n называется *степенью отображения*. Таким образом, степень введенного нами ранее отображения

$$f(\alpha) = n\alpha$$

равна n.

При непрерывной деформации отображения f значения $f(0°)$ и $f(360°)$ изменяются на одну и ту же величину. Следовательно, при непрерывных деформациях степень отображения сохраняется неизменной.

Если отображения f_0 и f_1 имеют равную степень n, то их можно непрерывно перевести друг в друга следующим образом. Каждому значению t, принадлежащему единичному интервалу (то есть удовлетворяющему неравенствам $0 \leqslant t \leqslant 1$), и любому α из C сопоставим отображение $f_t(\alpha)$, возникающего при равномерной деформации отображения $f_0(\alpha)$ по формуле

$$f_t(\alpha) = (1 - t) f_0(\alpha) + tf(\alpha).$$

Параметр t можно рассматривать как время, в течение которого происходит деформация. Проверим, удовлетворяет ли отображение $f_t(\alpha)$ предположению о том, что образы углов 0° и 360° [то есть $f(0°)$ и $f(360°)$] должны отличаться на величину $n \cdot 360°$:

$$\begin{aligned} f_t(360°) &= (1 - t) f_0(360°) + tf_1(360°) = \\ &= (1 - t) f_0(0°) + (1 - t) n \cdot 360° + tf_1(0°) + tn \cdot 360° = \\ &= (1 - t) f_0(0°) + tf_1(0°) + n \cdot 360° = \\ &= f_t(0°) + n \cdot 360°. \end{aligned}$$

Итак, при любом t образы $f_t(0°)$ и $f_t(360°)$ соответствуют одной и той же точке окружности C.

(Заметим, что если отображение f_1 определено соотношением $f_1(\alpha) = n\alpha$, то описанная нами непрерывная деформация сводится к разглаживанию складок.)

Итак, мы доказали следующее утверждение: *два отображения окружности C на себя принадлежат к одному классу в том и только в том случае, если они обладают одинаковыми степенями.*

В частности, любое отображение окружности C можно непрерывно деформировать в отображение, переводящее всю окружность в одну-единственную точку лишь тогда, когда его степень равна 0°.

Другое определение степени отображения

Можно ли аналогичным образом определить непрерывные отображения поверхности сферы S на себя? Оказывается, можно. Необходимо лишь иначе определить степень отображения.

Определение степени отображения окружности C на себя основано на том, что, введя градусную сетку на окружности, мы можем развернуть ее на прямой и построить график отображения. Проделать это на поверхности сферы нам не удастся.

Однако ничто не мешает нам ввести другое (эквивалентное первому) определение степени отображения окружности C на себя, которое допускает обобщение на случай поверхности сферы S.

Пусть g — отображение окружности C на себя, задаваемое соотношением

$$g(\alpha) = n\alpha \quad (n \text{ — целое число}).$$

Точки окружности, соответствующие углам α, значения которых отличаются на $p/|n| \cdot 360°$ ($p = 0, 1, \ldots$ $\ldots, |n| - 1$), при отображении g переходят в одну и ту же точку. Образ окружности C покрывает окружность C ровно $|n|$ раз. Так будет не при любом отображении степени n: если отображение имеет складку, то некоторые точки окружности покрыты более n раз. Это замечание показывает также, как нам надлежит поступать при подсчете числа прообразов данной точки. Каждому прообразу удобно приписать знак: положительный, если прообраз был покрыт при обходе окружности в положительном направлении (против часовой стрелки), и отрицательный, если, покрывая точку, мы двигались в отрицательном направлении. Как это выглядит на графике отображения, показано на рис. 46. Чтобы найти прообразы угла 180°, мы провели горизонтальные прямые, соответствующие $180° + m \cdot 360°$ (m — целое число), и нашли их точки пересечения с графиком $f(\alpha)$. Соответствующие значения служат решениями уравнения

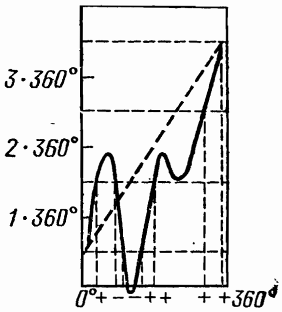


Рис. 46.

$$f(\alpha) = 180° + m \cdot 360°.$$

На рис. 46 их восемь. Точки пересечения, в которых кривая возрастает, следует считать положительными, точки пересечения, в которых кривая убывает, — отрицательными. Точки, в которых кривая $f(\alpha)$ касается проведенных нами горизонтальных прямых $180° + m \times 360°$, при подсчете числа прообразов не учитываются. Если подсчитать число прообразов угла 180° по этому правилу, то получится $5 - 2 + 0 = 3$ прообраза — ровно столько, сколько допускает степень отображения.

Применив это правило к образу любой точки при отображении

$$g(\alpha) = n\alpha \quad (n \text{ — целое число}),$$

мы обнаружим, что число прообразов в точности равно n (а не $|n|$). Если непрерывно изменять отображение, то у некоторых точек прообразы могут появляться и исчезать, но всегда лишь парами, причем один из прообразов пары соответствует положительной точке пересечения графика отображения с горизонтальной прямой, а другой — отрицательной точке пересечения (если исчезает или появляется лишь точка касания, то при подсчете числа прообразов она дает нулевой вклад). Таким образом, при непрерывной деформации отображения число прообразов любой точки сохраняется неизменным. Поскольку все непрерывные отображения степени n окружности C на себя можно получить непрерывной деформацией из отображения

$$g(\alpha) = n\alpha \quad (n \text{ — целое число}),$$

то справедливо следующее утверждение:

степень отображения равна числу прообразов любой точки, если при подсчете прообразов мы будем учитывать их знаки.

Наше определение степени отображения еще не вполне строго. Нетрудно представить себе отображения, у которых имеется бесконечно много складок, располагающихся все более тесно. При таком отображении у точки может быть бесконечно много прообразов. Как производить подсчет, если плюсов и минусов бесконечно много, никому не известно. Обычно поступают следующим образом. Отображение f заменяют сначала мало отличающимся от него отображением f', график которого кусочно линеен, и подсчитывают число прообразов какой-нибудь точки при отображении f'

(ставшее уже конечным). Затем проверяют, не зависит ли вычисленная степень отображения от выбора приближения, то есть остается ли она неизменной при замене отображения f' на эквивалентное ему отображение f''. Мы ограничимся сделанными нами замечаниями и не будем входить здесь в более тонкие детали подсчета степени отображения с бесконечным числом сужающихся складок.

Отображения сферы на себя

Теперь мы располагаем всем необходимым, чтобы приступить к рассмотрению отображений f сферы S на себя. Нам будет удобнее представлять себе поверхность S не в виде искривленной поверхности сферы, а в виде поверхности какого-нибудь многогранника, например тетраэдра. Грани тетраэдра мы разобьем на более мелкие треугольники.

Зададим по поверхности S ориентацию (стоя на ориентированной поверхности S, мы сможем отличить направление «налево» от направления «направо»). Каждый из мелких треугольников, например треугольник abc, получит при этом определенное направление обхода: либо $abca$, либо $acba$.

Относительно отображения f предположим сначала, что оно переводит треугольники одного разбиения поверхности S в треугольники, вообще говоря, другого ее разбиения, причем сужение отображения f на каждый треугольник в отдельности должно быть *барицентрическим* (см. также главу 6). Последнее означает, что если $a'b'c'$ — образ треугольника abc при отображении f, то центр тяжести треугольника $a'b'c'$ совпадает с образом центра тяжести треугольника abc. Подсчитаем число прообразов каждого треугольника-образа, приписав прообразам определенные знаки: будем считать прообраз abc треугольника $a'b'c'$ положительным, если ориентация треугольников abc и $a'b'c'$ совпадают, и отрицательным, если отображение f изменяет ориентацию треугольника abc на противоположную.

Нетрудно понять, что при отображении f все треугольника-образы имеют одинаковое число прообразов (если при подсчете прообразов придерживаться введенного нами правила знаков). Если мы вздумаем протащить какой-нибудь треугольник под складкой отображения, то число его прообразов может возрастать или уменьшаться, при этом всегда положительные и отрицательные прообразы будут возникать или исчезать поровну.

Сумму прообразов любого треугольника, взятых с их знаками, мы снова назовем степенью отображения f. Пока степень отображения определена лишь для отображения f, переводящего треугольники одного разбиения поверхности S в треугольники, вообще говоря, другого разбиения той же поверхности. Чтобы определить степень отображения для произвольного непрерывного отображения f, необходимо, как и в случае отображений окружности на себя, попытаться приближенно заменить f другим отображением f' и показать, что любое эквивалентное приближение приводит к одному и тому же значению степени отображения.

Наконец, так же как и в случае отображений окружности на себя, можно показать, что два отображения одного и того же класса (то есть два отображения, переводимых друг в друга непрерывной деформацией) обладают одной и той же степенью и, наоборот, два отображения с одинаковыми степенями принадлежат одному классу.

Пусть теперь S снова означает сферу. Тождественное отображение, оставляющее на месте все точки сферы, имеет степень 1. Отображение, ставящее в соответствие каждой точке сферы ее антипод — диаметрально противоположную точку, имеет степень -1, поскольку изменяет ориентацию сферы S. Существует лишь одно отображение степени 0. Оно переводит всю сферу S в какую-то одну точку, также принадлежащую S. Отображение степени n можно получить по следующему рецепту. Представим себе, что S — поверхность земного шара, и переведем любую ее точку в точку с той же широтой и n-кратной долготой. Экватор и все параллели отобразятся на себя, причем так, что образ любой параллели покроет ее n раз. При таком отображении каждая точка земной поверхности (кроме Северного и Южного полюсов) имеет в точности n прообразов (все прообразы положительны). Образ сферы S покрывает ее n раз.

При непрерывных преобразованиях степень отображения остается неизменной. В частности, тождественное отображение (имеющее степень 1) невозможно перевести непрерывным преобразованием в отображение, при котором образ сферы не покроет какую-нибудь точку сферы S, поскольку степень отображения, оставляющего некоторые точки сферы без крова,

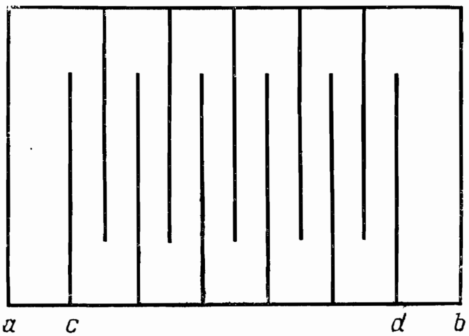


Рис. 47.

равна 0. Наглядно это можно представить себе следующим образом. Деревянный шар покрыли тонкой резиновой оболочкой. Невозможно добраться до поверхности деревянного шара, не разрывая, а лишь складывая и растягивая резиновую оболочку.

Подобное утверждение кажется очевидным. Сколь глубоко может лежать то, что представляется разумеющимся самим собой, если мы попытаемся подкрепить интуитивные представления строгими математическими доказательствами, видно из предыдущих рассуждений.

В топологии нередко случается, что утверждение, кажущееся на первый взгляд самоочевидным, при более внимательном рассмотрении оказывается неверным.

Можно ли вырезать в листе бумаги размером со страницу лежащей перед вами книги дыру, через которую пролезет взрослый человек?

Конечно, вы правы: эту задачу решить совсем просто. Сложим лист бумаги пополам (сложенный лист бумаги изображен на рис. 47, ab — линию сгиба), разрежем сначала вдоль линий, образующих на рис. 47 две гребенки, а затем вдоль линии сгиба от c до d. Развернув остатки листа, мы обнаружим в нем дыру нужных размеров.

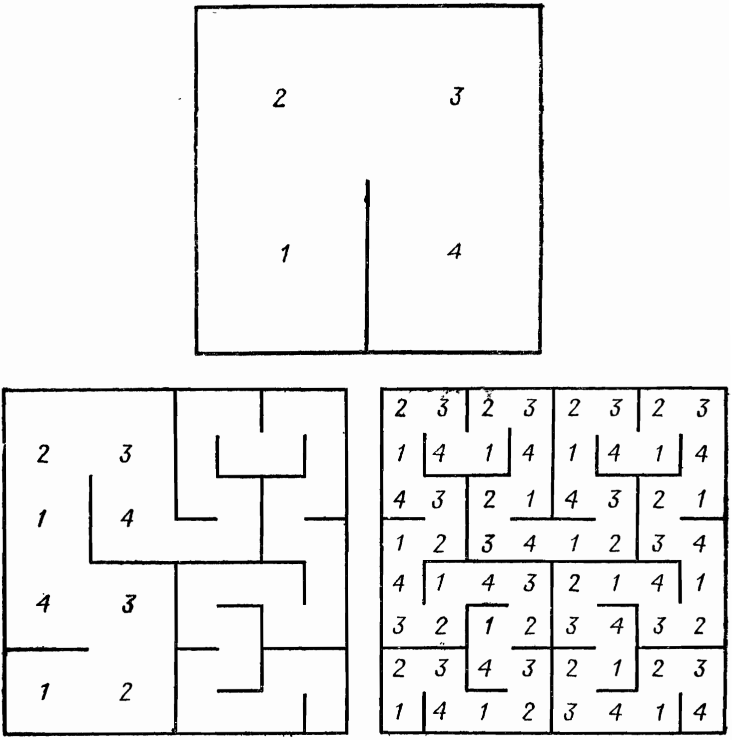


Рис. 48.

В листке бумаги размером с почтовую марку также можно прорезать дыру, через которую пройдет человек, только разрезов придется провести побольше.

Для нас в этой задачке особенно важно, что поверхность можно рассечь на части, довольно близко напоминающие линии. Впрочем, в следующем примере столь важное свойство поверхностей выступает еще ярче.

Разложим квадрат (изображенный на рис. 48 сверху) на четыре квадрата 1, 2, 3, 4 со сторонами, равными половине стороны исходного квадрата. Проведем разрез вдоль общей стороны квадратов 1 и 4. Каждый из четырех получившихся квадратов в свою очередь разложим на четыре квадрата с вдвое меньшими сторонами, перенумеруем их так, как показано на рис. 48 слева, и проведем разрезы вдоль общих сторон квадратов с номерами 1 и 4 (на левой половинке необходимо провести все недостающие разрезы). Третий этап разрезания изображен на рис. 48 справа. Как и прежде, каждый из квадратов разделен на 4 квадрата с вдвое меньшими сторонами, «четвертушки» перенумерованы и проведены разрезы вдоль общих сторон квадратов 1 и 4. Аналогичным образом мы поступим и на четвертом этапе. Разрезание квадрата

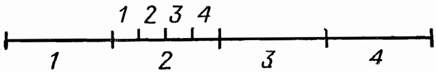


Рис. 49.

можно мысленно продолжать до бесконечности, проводя пятые, шестые, седьмые и так далее разрезы. Чем дальше мы будем продвигаться, тем длиннее и уже будет становиться лабиринт. Нам удастся построить его потому, что на каждом этапе четвертушки квадрата мы нумеруем особым образом, следя за тем, чтобы четвертый подквадрат в квадрате с номером a имел общую сторону с первым подквадратом в квадрате с номером $a+1$.

Предположим, что мы продолжили разрезание квадрата до бесконечности и рассекли поверхность квадрата, превратив ее в некоторую кривую линию. Иначе говоря, мы отобразили квадрат на какую-то кривую. Наше отображение не может быть непрерывным: проводя разрезы, мы из одной точки «делали» две или четыре. Более точно то, что происходит при разрезании, можно сформулировать следующим образом.

Сопоставим исходному квадрату отрезок (рис. 49). При первом разрезании разделим этот отрезок на четыре равные части 1, 2, 3, 4. При втором снова разделим на четыре равные части 1, 2, 3, 4 каждую из четвертушек и будем продолжать так до бесконечности.

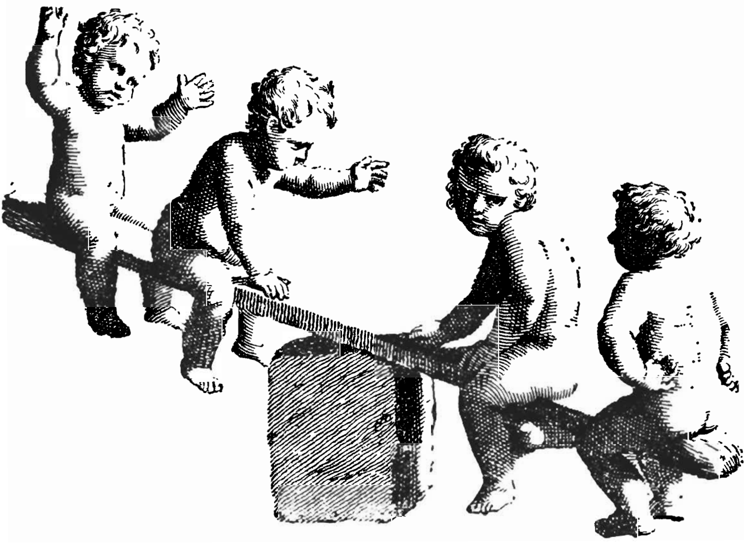
Задавать положение точки x на исходном отрезке мы можем, указывая на каждом этапе разрезания номер того подотрезка, на котором она находится. Выписав подряд все адреса точки x, мы получим бесконечную последовательность, например такую: 324133144... . Некоторые точки исходного отрезка порождают не одну, а две такие последовательности. Например, середину отрезка можно представить и в виде последовательности 2444..., и в виде последовательности 31111... . Совершенно аналогично можно задать положение точки в квадрате: для этого необходимо лишь при каждом разрезании отметить номер квадрата, в котором находится интересующая нас точка.

Сопоставим теперь точку отрезка и точку квадрата, порождающие одну и ту же числовую последовательность. У нас получится непрерывное отображение отрезка на квадрат: образ отрезка построенного нами отображения заполнит весь квадрат, причем близким точкам отрезка будут соответствовать близкие точки квадрата. Это непрерывное отображение представляет собой не что иное, как отображение, обратное разрезанию, с которого мы начали. Из отрезка нам удалось склеить квадрат.

Часть открытия столь необычной кривой, полностью покрывающей квадрат, принадлежит итальянскому математику Дж. Пеано (1890).

Глава 6

«Дайте мне точку опоры, и я переверну Землю»



Архимедовы чудеса

Предания до наших дней сохранили сведения о всякого рода технических чудесах, которыми Архимед приводил в изумление не только жителей Сиракуз, но и римлян, осаждавших его родной город. Среди прочих рассказов до нас дошли упоминания об устройстве, при помощи которого один человек, стоя на берегу, мог поднять из моря тяжело груженный корабль. Утверждают, будто, объясняя действие своего подъемного устройства, великий математик древности сказал царю Генону: «Дайте мне точку опоры,

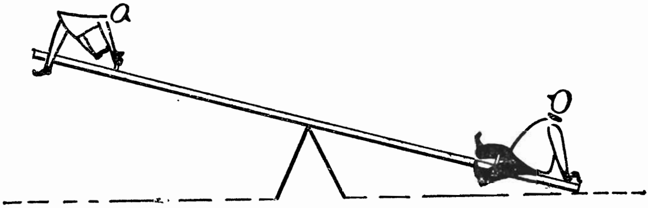


Рис. 50.

и я переверну Землю». По другому преданию, эти гордые слова относились к правилу рычага, которому Архимед посвятил одно из своих сочинений. Рычаг действительно позволяет поднимать большие грузы малыми силами.

В этой главе мы постепенно превратим теорию рычага в один из разделов современной математики.

Рычаг

Два мальчугана садятся на качели (рис. 50). Доска опускается в сторону того из них, чей вес больше, и остается в равновесии лишь в том случае, если оба мальчика весят одинаково.

А могут ли качели остаться в равновесии, несмотря на разницу в весе мальчиков? Конечно могут, только тот из мальчиков, чей вес больше, должен сместиться к середине доски, то есть к точке опоры. Чем ближе груз к середине доски, тем меньше он тянет доску в свою сторону, как бы становясь легче. Можно поступить и иначе: если тот из мальчиков, чей вес меньше, сядет подальше от точки опоры, то он сильнее потянет доску в свою сторону.

Спортсмен собирается прыгнуть в воду с трамплина (рис. 51). Чем дальше он отходит от края площадки, тем больше прогибается доска. Вес прыгуна

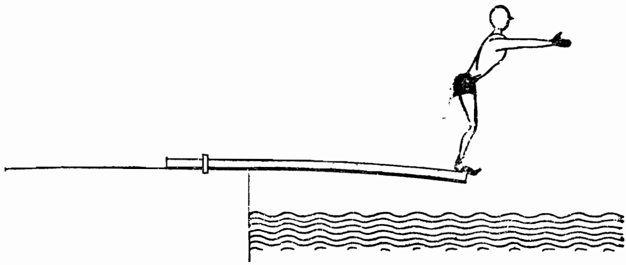


Рис. 51.

сказывается на прогибе тем сильнее, чем дальше он отстоит от места закрепления доски.

Два носильщика, шагая друг за другом, несут на длинном шесте корзину (рис. 52). Если корзина висит посредине шеста, то им обоим приходится прикладывать к шесту одинаковую силу. Если корзина переместится поближе к одному из концов шеста, то тот из носильщиков, кто окажется дальше от корзины, как

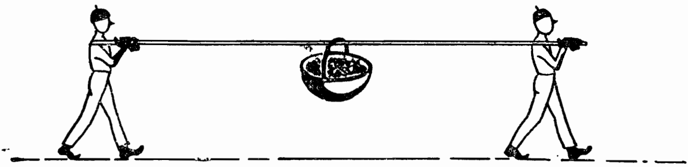


Рис. 52.

бы переложит часть своего груза на плечи другого. Сила, которую каждый из носильщиков прикладывает к своему концу шеста, тем больше, чем ближе к этому концу корзина.

Если к правому и левому плечу весов подвешены одинаковые грузы, то коромысло весов находится в равновесии, по крайней мере коромысло равноплечих весов, изображенных на рис. 53. Само коромысло весов раз и навсегда условимся считать «невесомым», то есть пренебрежем его собственным весом. На неравноплечих весах картина будет иная. Такие весы весьма схематично изображены на рис. 54: их левое плечо OA вдвое длиннее правого плеча OB. Коромысло весов AB может поворачиваться вокруг центра

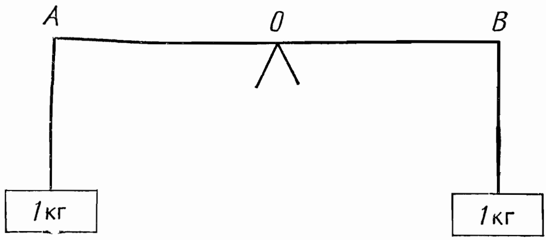


Рис. 53.

O — точки опоры. Чтобы коромысло весов находилось в равновесии, груз, подвешенный в точке B, должен быть вдвое больше груза, подвешенного в точке A. В общем случае, если мы хотим, чтобы коромысло весов с отношением плеч

$$OA:OB=a:b$$

находилось в равновесии, то груз, подвешенный в точке A, должен относиться к грузу, подвешенному в точке B, как $b:a$ (рис. 55).

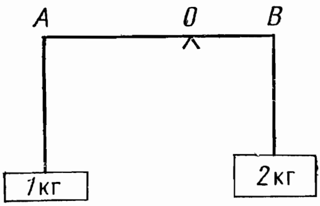


Рис. 54.

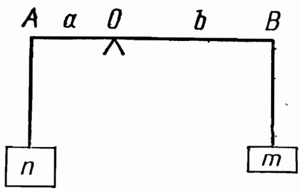


Рис. 55.

То же самое можно выразить иначе. Груз в m кГ — это сила в m кГ. Если линия действия этой силы проходит на расстоянии a от точки O, то силе m соответствует момент ma, равный произведению силы на плечо a. Аналогичным образом силе в n кГ, линия действия которой отстоит от точки O на расстояние b, мы сопоставляем момент nb.

Если удвоить груз, оставив без изменения плечо, то момент возрастает вдвое. То же самое произойдет, если оставить без изменения груз и увеличить вдвое длину плеча относительно точки O. Если одновременно удвоить груз и сократить наполовину плечо, то момент относительно точки O останется без изменения. Момент сохранится и в том случае, если мы уменьшим вдвое груз, удлинив плечо в два раза.

Равные грузы m на равноплечих весах с плечом a создают справа и слева от точки опоры O моменты

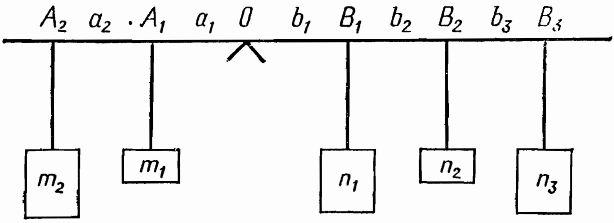


Рис. 56.

ma. Поскольку моменты равны по величине, коромысло весов находится в равновесии. Если взять неравноплечие весы с левым плечом a и правым плечом b и подвесить в точках A и B грузы m и n, то слева от точки опоры O мы получим момент ma (равный произведению веса m на плечо a), а справа момент nb. Чтобы коромысло весов находилось в равновесии, оба момента должны быть равны:

$$ma = nb.$$

Это равенство можно также представить в следующем виде:

$$m : n = b : a,$$

иначе говоря, грузы должны изменяться обратно пропорционально длине плеча.

Теперь представим себе коромысло весов с опорой в точке O (рис. 56), к которому в точках A_1, A_2, ..., находящихся на расстояниях a_1, a_2, ... от O, подвешены грузы m_1, m_2, ..., а в точках B_1, B_2, ..., находящихся от точки опоры на расстояниях b_1, b_2, ..., — грузы n_1, n_2, ...

Выпишем моменты, образуемые всеми грузами, и сложим их. Полный момент всех грузов, приложенных слева от точки опоры O, равен

$$m_1a_1 + m_2a_2 + \ldots,$$

полный момент всех грузов, приложенных слева от точки O, равен

$$n_1b_1 + n_2b_2 + \ldots .$$

Равновесие наступит в том случае, если

$$m_1a_1 + m_2a_2 + \ldots = n_1b_1 + n_2b_2 + \ldots .$$

Если в некоторой точке, например в точке A, находящейся на расстоянии a от точки опоры O, подвесить два груза m и m', то суммарный момент, создаваемый этими грузами относительно точки O, равен

$$ma + m'a.$$

Но вместо двух грузов в той же точке A можно подвесить один груз $m + m'$, создающий момент относительно точки O, равный

$$(m + m')a.$$

Оба момента должны быть равны, и действительно, как известно из алгебры,

$$(m + m')a = ma + m'a.$$

Нашим выводам можно придать более изящный вид, если мы условимся иначе записывать моменты. Груз, подвешенный к коромыслу весов справа от точки опоры O, стремится повернуть коромысло вправо относительно наблюдателя. Груз, подвешенный к коромыслу весов слева от точки опоры O, стремится повернуть его влево. Будем считать моменты положительными или отрицательными в зависимости от того, в какую сторону они стремятся повернуть коромысло: вправо или влево. Знак момента можно определить следующим образом. Будем считать расстояния вдоль коромысла весов вправо (от точки опоры O) положительными, а влево — отрицательными. Умножив соответствующие расстояния на величины грузов, получим моменты сил. Если один из сомножителей в произведении изменяет знак на противоположный, то и само произведение изменяет знак на противоположный.

Отсюда следует, что по новому правилу моменты всех грузов, находящихся слева от точки опоры O, отрицательны.

Следуя старому правилу, мы формулировали условия равновесия коромысла весов так: полный момент, создаваемый грузами слева от точки опоры, должен быть равен полному моменту, создаваемому грузами справа от точки опоры. Новое правило позволяет нам сформулировать условие равновесия еще проще: *полный момент, создаваемый всеми грузами, должен быть равен нулю.*

И хотя теперь мы просто складываем *все* моменты, те из них, которые стремятся повернуть коромысло весов влево, автоматически получаются отрицательными.

Поставим вопрос несколько иначе. Предположим, что у нас есть коромысло весов и заданные грузы. Как следует их повесить, чтобы коромысло находилось в равновесии?

Прежде всего выберем произвольно начало отсчета — точку O. Пусть a_1, a_2, ... — расстояния, на которых находятся от нее точки подвеса грузов m_1, m_2, ... (расстояния считаются положительными, если точки подвеса лежат справа от O, и отрицательными, если точки подвеса лежат слева от O). Относительно точки O полный момент, создаваемый грузами m_1, m_2, ..., равен

$$M = m_1a_1 + m_2a_2 + \ldots .$$

Теперь сдвинем начало отсчета на расстояние x, тогда расстояния от нового начала до точек подвеса грузов станут равными $a_1 - x$, $a_2 - x$, ..., а полный момент

$$m_1(a_1 - x) + m_2(a_2 - x_2) + \ldots = \\ = (m_1a_1 + m_2a_2 + \ldots) - (m_1 + m_2 + \ldots)x = \\ = M - mx,$$

где $m = m_1 + m_2 + \ldots$ — суммарный груз. Если, совместив точку опоры с новым началом отсчета, мы хотим получить равновесие, то новое значение полного момента должно равняться нулю, откуда $M = mx$ и

$$x = \frac{M}{m}.$$

Точка, в которой можно подпереть коромысло весов так, чтобы оно находилось в равновесии, называется *центром тяжести*. Как мы только что показали, *положение центра тяжести можно найти, разделив полный момент, создаваемый всеми грузами, на суммарный груз*.

Точку, относительно которой мы вычисляем полный момент и находим положение центра тяжести, можно выбирать произвольно. Действительно, при сдвиге начала отсчета на c расстояния a_1, a_2, ... переходят в $a_1 - c$, $a_2 - c$, ..., полный момент M становится равным $M - mc$ и отношение M/m, из которого мы определяем координату центра тяжести, также уменьшится на c.

Вычеркнем теперь некоторые грузы, например грузы m_1, m_2, ..., подвешенные на расстояниях a_1, a_2, ... от центра тяжести (кроме них мы можем вычеркнуть также грузы весом n_1, n_2, ..., подвешенные на расстояниях b_1, b_2, ... от центра тяжести), поскольку эти грузы пока не будут интересовать нас, и заменим их одним грузом $m_1 + m_2 + \ldots$. Этот груз подвесим в центре тяжести вычеркнутых грузов. Мы утверждаем, что положение центра тяжести всей системы от такого преобразования не изменится.

Действительно, вычеркнутые грузы вносят в полный момент вклад, равный $m_1a_1 + m_2a_2 + \ldots$. Новый груз $m_1 + m_2 + \ldots$ вносит в полный момент вклад $(m_1 + m_2 + \ldots)x$, где x — координата центра тяжести вычеркнутых грузов. Но

$$x = \frac{m_1a_1 + m_2a_2 + \ldots}{m_1 + m_2 + \ldots},$$

а это означает, что новый груз вносит в полный момент в точности такой же вклад, какой вносили вычеркнутые грузы. Следовательно, положение центра тяжести всей системы остается неизменным.

Итак, *любую систему грузов можно заменить суммарным грузом, сосредоточенным в центре тяжести системы.*

Продвинемся теперь еще на один шаг: введем отрицательные грузы. Представим себе коромысло весов, нагруженное так, как было показано на рис. 56. Что следует понимать под отрицательным грузом?

Очевидно, груз, который тянет не вниз, а вверх. Нечто подобное мы можем создать, изменив направление силы, с которой действует груз на плечо коромысла, например, при помощи блока (на рис. 57 к правому плечу коромысла приложена сила A, равная m кГ и направленная вверх). Ясно, что на коромысла отрицательный груз m кГ, сосредоточенный в точке A, оказывает такое же действие, как положительный груз m кГ, сосредоточенный в точке B, если точка A лежит слева от точки опоры O на таком же расстоянии, на каком точка B лежит справа от O. Вычисляя моменты, создаваемые тем и другим грузом, получаем: груз $-m$, действуя на плечо a, создает момент

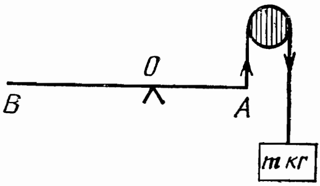


Рис. 57.

$$(-m)\,a,$$

груз m, действуя на плечо $-a$, создает момент

$$m\,(-a).$$

По правилам алгебры оба момента равны

$$-ma.$$

Любопытно заметить, что отрицательный груз в точке B оказывает на коромысло весов такое же действие, как соответствующий положительный груз в точке A. В первом случае груз вносит в полный момент системы вклад

$$(-m)(-a),$$

во втором —

$$m \cdot a.$$

Согласно алгебраическим правилам, эти два момента также равны.

Итак, мы получаем возможность прикладывать к коромыслу весов любые положительные и отрицательные грузы и вычислять создаваемый ими полный момент. Зная суммарный груз m и полный момент M

системы, мы находим положение центра тяжести грузов из соотношения

$$x = \frac{M}{m}.$$

Но на этот раз необходимо действовать осторожно. Поскольку грузы могут быть положительными и отрицательными, суммарный груз может оказаться равным 0, а делить на нуль мы не умеем, да и делать этого нельзя. Таким образом, в некоторых случаях центра тяжести системы грузов не существует. Например, если подвесить к коромыслу весов в двух различных точках грузы в 1 кГ, один — как обычно, а другой — при помощи перекинутой через блок нити, то коромысло весов, где бы мы его ни подпирали, никогда не придет в равновесие.

Плоскость

Рассмотрим теперь геометрическую плоскость, то есть невесомую горизонтальную пластинку, на которой расставлены грузы. Попытаемся найти точку, в которой можно было бы подпереть пластинку так, чтобы она находилась в состоянии равновесия (такая точка называется *центром тяжести*).

Предположим, например, что три груза по 1 кГ каждый расставлены в вершинах треугольника ABC (рис. 58). Из того, что нам уже известно, следует, что грузы в 1 кГ, расположенные в точках A и B, можно заменить грузом в 2 кГ, поместив его в середину F отрезка AB. Итак, вместо исходной системы трех грузов в 1 кГ, размещенных в вершинах треугольника ABC, можно рассматривать эквивалентную систему двух грузов: груз в 1 кГ в точке C и груз в 2 кГ в точке F. Новая система грузов обладает центром тяжести S, расположенным между крайними точками отрезка CF. Полный момент системы двух грузов относительно точки S равен 0, поэтому точка S делит отрезок CF в отношении, обратном отношению грузов в точках F и C:

$$CS : SF = 2 : 1.$$

Таким образом, точка S совпадает с точкой, называемой в геометрии центром тяжести треугольника ABC.

Если бы мы сначала заменили грузы в точках A и C удвоенным грузом в середине отрезка E, то в конце концов пришли бы к той же самой точке S. Как известно, медианы треугольника пересекаются в его центре тяжести S, который делит любую из медиан в отношении $2:1$ (считая от соответствующей вершины).

От частного случая — трех грузов, расставленных в вершинах треугольника, — перейдем к рассмотрению общего случая: определению центра тяжести системы

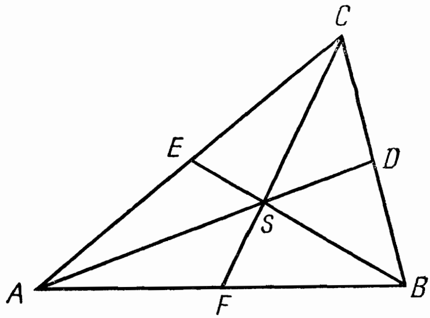


Рис. 58.

грузов, расставленных на плоской невесомой пластинке. Воспользуемся уже испытанным приемом. Заменим два груза их суммой и поместим ее в центр тяжести двух грузов-слагаемых. Продолжая присоединять по одному грузу, мы в конце концов сольем все грузы в один. Точка плоскости, в которой будет помещен этот груз, и будет центром тяжести исходной системы грузов.

Возникает вопрос: однозначно ли определен центр тяжести? Получим ли мы ту же самую точку, если будем объединять исходные грузы в другой последовательности?

В действительности положение центра тяжести определено однозначно (по крайней мере в тех случаях, когда сумма весов отлична от нуля). Чтобы доказать это утверждение, обратимся вновь к моментам. Прежде чем определять, что такое момент силы относительно точки, определим момент силы относительно прямой, называемой также осью.

Проведем на плоскости прямую X и условимся называть полуплоскость, расположенную по одну сторону от нее, правой, а полуплоскость, расположенную по другую сторону от нее, левой. Расстояние от прямой, или оси, X до точки (длину перпендикуляра, опущенного из точки на ось X) будем считать положительным для точек правой полуплоскости и отрицательным для точек левой полуплоскости. Положительные грузы, расположенные по разные стороны

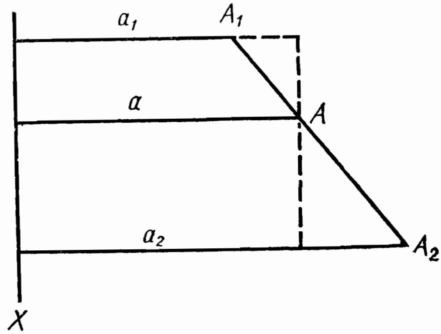


Рис. 59.

от оси, стремятся повернуть плоскость в противоположных направлениях. Груз m, находящийся на (положительном или отрицательном) расстоянии a от оси X, создает относительно нее момент ma. Суммируя моменты, создаваемые отдельными грузами, мы можем вычислить полный момент данной системы грузов относительно оси X.

Что произойдет с полным моментом, если два груза m_1 и m_2, расположенные в точках A_1 и A_2, заменить грузом $m_1 + m_2$, поместив его в центр тяжести этих грузов (рис. 59)?

Точка A делит отрезок A_1A_2 в отношении $m_2 : m_1$. Точки A_1, A_2, A находятся от оси X на расстояниях a_1, a_2, a. Проведя через точку A прямую, параллельную оси X, нетрудно убедиться в том, что $(a - a_1) : (a_2 - a) = m_2 : m_1$. Следовательно,

$$m_1 (a - a_1) = m_2 (a_2 - a),$$

$$(m_1 + m_2) a = m_1 a_1 + m_2 a_2.$$

В правой части последнего равенства стоит полный момент грузов m_1 и m_2, расположенных в точках A_1 и A_2, в левой части — момент суммарного груза, находящегося в центре тяжести грузов m_1 и m_2. Итак, мы доказали, что

полный момент относительно оси X не изменится, если вместо двух грузов взять груз, равный их сумме, и поместить его в центр тяжести двух исходных грузов.

Таким образом, на вопрос о том, зависит ли положение центра тяжести системы грузов от последовательности, в которой мы объединяем их в один груз,

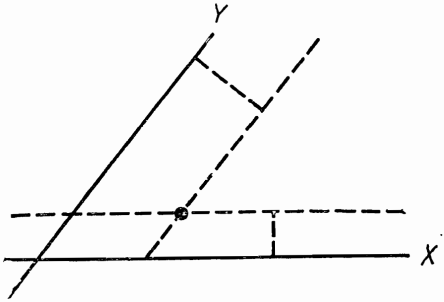


Рис. 60.

нам удалось получить ответ, хотя и не полный. Во всяком случае момент центра тяжести относительно оси X определяется однозначно: он совпадает с полным моментом заданной системы грузов. Но тем самым однозначно определено и расстояние от центра тяжести системы до оси X: его мы найдем, разделив полный момент на суммарный груз (если он отличен от 0).

Ось X на плоскости можно выбирать совершенно произвольно. Тем самым доказано, что расстояние от центра тяжести системы грузов до любой прямой на плоскости однозначно определено. Следовательно, и положение самого центра тяжести на плоскости также однозначно определено.

Впрочем, для того чтобы однозначно задать на плоскости положение точки, достаточно знать расстояния от нее до двух любых пересекающихся прямых X и Y (рис. 60): нужно лишь провести прямые, параллельные прямым X и Y и отстоящие от них на заданные (положительные или отрицательные) расстояния. Интересующая нас точка совпадает с точкой пересечения проведенных прямых.

Еще одно замечание. Если суммарный груз равен 0, то ожидать, что нам удастся найти центр тяжести системы грузов, не приходится. Но и при отличной от 0 сумме всех грузов может случиться, что при построении центра тяжести системы мы столкнемся на одном из промежуточных этапов с нулевыми грузами. Если грузы m_1 и m_2 равны по величине, но имеют противоположные знаки, то построение центра тяжести всей системы следует начинать не с них. Поскольку

$$m_1 + m_2 + m_3 + \ldots \neq 0,$$

то среди грузов, образующих систему, непременно должна найтись пара таких, что их сумма отлична от 0. Это же утверждение остается в силе и для любого последующего этапа построения центра тяжести всей системы.

Барицентрические координаты

Пусть A и B — две различные точки плокости. Если в них поместить грузы m и n с отличной от 0 суммой, то центр тяжести этих точек Q будет лежать на прямой AB, причем $m \cdot \overrightarrow{QA} + nQB = 0$, или в другой формулировке

$$\overrightarrow{AQ} : \overrightarrow{QB} = n : m.$$

И наоборот, если требуется, чтобы центр тяжести двух грузов совпадал с точкой Q, то нужно лишь взять два груза n и m, отношение которых равно $\overrightarrow{AQ} : QB$. В частности, при $n = 0$ центр тяжести совпадает с точкой A, а при $m = 0$ — с точкой B.

Сами по себе грузы n и m не существенны, важно лишь их отношение $n : m$. Если m и n умножить на одно и то же число, то положение центра тяжести от этого не изменится. Центр тяжести располагается *между* точками A и B, если грузы положительны. Если грузы различных знаков, то центр тяжести лежит вне отрезка AB и его продолжении со стороны большего по абсолютной величине груза.

Итак, любую точку прямой AB можно получить, задавая отношение грузов $m : n$.

Выберем на плоскости три точки A_1, A_2, A_3, не лежащие на одной прямой. Грузами m_1, m_2, m_3 с суммой, отличной от 0, расположенным в точках A_1, A_2, A_3, соответствует центр тяжести P. Положение его также зависит лишь от отношений грузов m_1, m_2, m_3. Наоборот, располагая в точках A_1, A_2, A_3 надлежащим

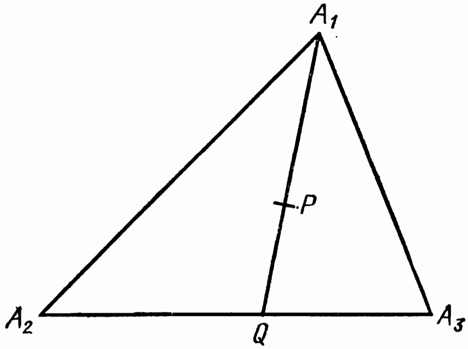


Рис. 61.

образом подобранные грузы, можно совместить их центр тяжести P с любой заранее заданной точкой плоскости. Сделать это можно следующим образом. Если заданная точка плоскости P совпадает с точкой A_1, то $m_1 = 1$, $m_2 = m_3 = 0$. Нетрудно видеть, что центр тяжести такой «системы» грузов действительно совпадает с точкой A_1. Аналогичным образом поступим, если $P = A_2$ или $P = A_3$.

Предположим, что точка P не совпадает ни с одной из точек A_1, A_2, A_3. Отрезок прямой, соединяющий точку P с одной из вершин треугольника $A_1A_2A_3$, может оказаться параллельным противолежащей стороне, но все три отрезка не могут быть параллельны противолежащим сторонам одновременно. Следовательно, по крайней мере одна прямая PA_i пересекает противолежащую сторону A_jA_k $(j \neq i,\ k \neq i)$. Предположим, что прямая PA_1 (рис. 61) пересекает сторону A_2A_3 в точке Q (в остальных случаях нам пришлось бы провести аналогичное построение). Поместим в вершины A_2 и A_3 грузы m_2 и m_3, центр тяжести которых совпадает с точкой Q. Как подобрать такие грузы, нам уже известно. «Сольем» грузы m_2 и m_3 и заменим их грузом $m_2 + m_3$, поместив его в точку Q. В вершине A_1 расположим груз m_1, подобрав его так, чтобы центр тяжести грузов m_1 в точке A_1 и $m_2 + m_3$ в точке Q совпадал с точкой P (как это делается, мы уже видели). Таким образом, точка P совпадает с центром тяжести

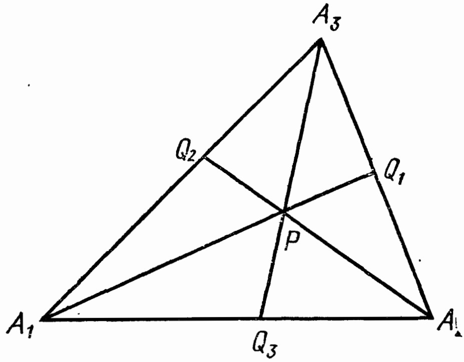


Рис. 62.

грузов m_1, m_2, m_3, расположенных в точках A_1, A_2, A_3, что и требовалось.

Располагая надлежащим образом подобранные грузы m_1, m_2, m_3 в заданных точках A_1, A_2, A_3, можно совместить центр тяжести этой системы с любой точкой плоскости. Положение центра тяжести зависит лишь от отношений грузов m_1, m_2, m_3, а не от их величин. Разделив величину каждого груза на их сумму (заведомо отличную от 0), мы тем самым нормируем их сумму к 1. В дальнейшем мы всегда будем предполагать, что $m_1 + m_2 + m_3 = 1$.

Три числа m_1, m_2, m_3, соответствующие точке плоскости P, называются ее *барицентрическими координатами* (координатами центра тяжести). Впервые барицентрические координаты ввел Мёбиус (1827).

Рассмотрим точку P с барицентрическими координатами m_1, m_2, m_3 и соединим ее прямыми с вершинами A_1, A_2, A_3 (рис. 62). Пусть Q_1, Q_2, Q_3 — точки пересечения прямых PA_i с противолежащими сторонами A_jA_k $(j \neq i,\ k \neq i)$.

Тогда

$$\frac{\overrightarrow{A_1Q_3}}{\overrightarrow{Q_3A_2}} = \frac{m_2}{m_3}, \quad \frac{\overrightarrow{A_2Q_1}}{\overrightarrow{Q_1A_3}} = \frac{m_3}{m_1}, \quad \frac{\overrightarrow{A_3Q_2}}{\overrightarrow{Q_2A_1}} = \frac{m_1}{m_2}.$$

Перемножив в отдельности левые и правые стороны всех трех равенств, получим

$$\overrightarrow{A_1Q_3} \cdot \overrightarrow{A_2Q_1} \cdot \overrightarrow{A_3Q_2} = \overrightarrow{Q_1A_3} \cdot \overrightarrow{Q_2A_3} \cdot \overrightarrow{Q_3A_2}.$$

Это соотношение известно под названием теоремы Менелая (в честь древнегреческого математика).

При $m_1 = 1$, $m_2 = m_3 = 0$ мы получаем точку A_1, при $m_2 = 1$, $m_1 = m_3 = 0$ — точку A_2, при $m_3 = 1$, $m_1 = m_2 = 0$ — точку A_3. Все точки с $m_1 = 0$ лежат на прямой A_2A_3, все точки с $m_2 = 0$ — на прямой A_1A_3, все точки с $m_3 = 0$ — на прямой A_1A_3.

Барицентрическая координата m_1 положительна по ту сторону от прямой A_2A_3, по которую расположена точка A_1, и отрицательна по другую сторону. Аналогичное утверждение справедливо и относительно барицентрических координат m_2 и m_3. Точки, у которых положительны все три барицентрические координаты, заполняют внутренность треугольника $A_1A_2A_3$.

При $P = A_1$ барицентрическая координата m_1 равна 1, а точки P с $m_1 = 0$ лежат на прямой A_2A_3. Где лежат точки с $m_1 = 1/2$? Если $m_1 = 1/2$, то $m_2 + m_3 = 1/2$. Следовательно, мы должны поместить одинаковые грузы в точке A_1 и в некоторой произвольно выбранной точке Q на прямую A_2A_3. Точка P описывает прямую, параллельную прямой A_2A_3, и делит пополам расстояние, отделяющее точку A_1 от прямой A_2A_3 (на рис. 63 показаны различные линии уровня барицентрической координаты m_1). Пользуясь чертежом, можно убедиться в том, что прямые, соответствующие различным значениям барицентрической координаты m_1, должны быть параллельными. Несколько ниже мы приведем общее доказательство этого утверждения.

Но пойдем дальше. Выясним, для каких точек P выполняется равенство $m_2 = m_3$. В точки A_2 и A_3 необходимо поместить равные грузы, их центр тяжести совпадает с серединой отрезка A_2A_3. Затем мы должны поместить какой-то груз в точку A_1. Центр тяжести системы трех грузов лежит на медиане, проведенной из вершины A_1 к стороне A_2A_3 (на рис. 64, кроме медианы, проведено еще несколько прямых, проходящих через вершину A_1).

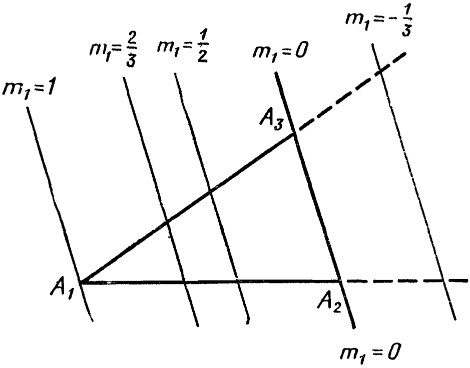


Рис. 63.

В общем случае можно составить уравнение

$$p_1m_1 + p_2m_2 + p_3m_3 = p_0, \qquad (*)$$

где p_0, p_1, p_2, p_3 — заданные числа, и спросить, какая точка P ему удовлетворяет, то есть барицентрические

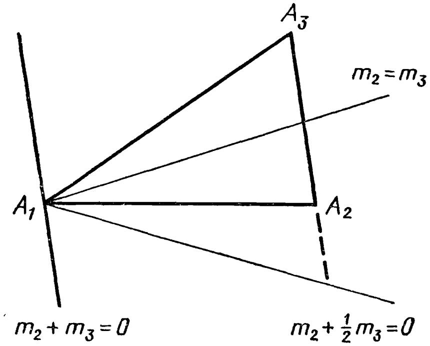


Рис. 64.

координаты m_1, m_2, m_3 какой точки P при подстановке их в уравнение $(*)$ обращают его в тождество.

Заметим, что для любой точки должно выполняться равенство

$$m_1 + m_2 + m_3 = 1. \qquad (**)$$

Вычитая из левой и правой частей уравнения (*) левую и правую части равенства (**), умноженные на p_0, преобразуем уравнение (*) к следующему виду:

$$(p_1 - p_0) m_1 + (p_2 - p_0) m_2 + (p_3 - p_0) m_3 = 0.$$

Итак, не ограничивая общности, мы можем впредь рассматривать лишь уравнения вида

$$q_1 m_1 + q_2 m_2 + q_3 m_3 = 0 \qquad (***)$$

с заданными коэффициентами q_1, q_2, q_3.

В общем случае существуют решения, удовлетворяющие одновременно уравнениям (**) и (***). Если эти уравнения противоречивы, то есть если $q_1 = q_2 = q_3 \neq 0$, то система уравнений (**) и (***) не допускает ни одного решения. В противном случае из одного уравнения следовало бы соотношение $m_1 + m_2 + m_3 = 1$, а из другого $m_1 + m_2 + m_3 = 0$. Если $q_1 = q_2 = q_3 = 0$, то уравнениям (**) и (***) удовлетворяют барицентрические координаты любой точки P.

Мы утверждаем, что если координаты точки P и P' удовлетворяют уравнениям (**) и (***), то координаты любой точки R на прямой PP' также удовлетворяют этим уравнениям. Действительно, точку R можно рассматривать как центр тяжести грузов α и β, расположенных в точках P и P'. Если P — центр тяжести грузов m_1, m_2, m_3, расположенных в точках A_1, A_2, A_3, а P' — центр тяжести грузов m_1', m_2', m_3', расположенных в тех же точках, то точку R можно считать центром тяжести грузов $\alpha m_1 + \beta m_1'$, $\alpha m_2 + \beta m_2'$, $\alpha m_3 + \beta m_3'$, расположенных в точках A_1, A_2, A_3. Но

$$q_1(\alpha m_1 + \beta m_1') + q_2(\alpha m_2 + \beta m_2') + q_3(\alpha m_3 + \beta m_3') = \alpha(q_1 m_1 + q_2 m_2 + q_3 m_3) + \beta(q_1 m_1' + q_2 m_2' + q_3 m_3') = 0,$$

поскольку координаты точек P и P' удовлетворяют уравнению (***). Следовательно, координаты точки R также удовлетворяют уравнению (***), что и требовалось доказать. [Если координаты точки R нормировать к 1, разделив каждую из них на их сумму, то «нормированные» координаты, помимо уравнения (***), будут удовлетворять и уравнению (**).]

Итак, вместе с координатами двух точек P и P' уравнению (***) удовлетворяют координаты всех точек прямой PP'. Если точки, координаты которых удовлетворяют уравнению (***), заполняют не одну, а несколько прямых, то они заполняют всю плоскость. Именно так и происходит при $q_1 = q_2 = q_3 = 0$. Если $q_1 = q_2 = q_3 \neq 0$, то решения не существует. Во всех

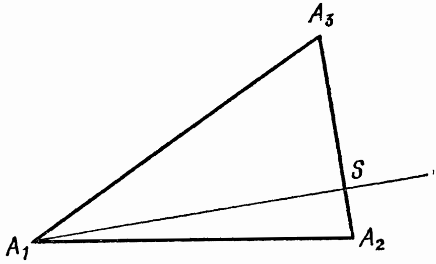


Рис. 65.

остальных случаях решение уравнения (***) существует, и точки с допустимыми значениями координат заполняют прямую.

Можно ли, решая уравнение (***), получить любую прямую? Ответ на этот вопрос утвердителен. Проведем прямую, проходящую через вершину A_1 (рис. 65). На этой прямой должна лежать точка с барицентрическими координатами $m_1 = 1$, $m_2 = 0$, $m_3 = 0$, удовлетворяющими уравнению (***). Следовательно, $q_1 = 0$. Таким образом, уравнение прямой, проходящей через точку A_1, должно иметь вид

$$q_2m_2 + q_3m_3 = 0.$$

Пусть S — точка пересечения проведенной прямой с прямой A_2A_3. Тогда

$$m_2 \cdot \overrightarrow{SA_2} + m_3 \cdot \overrightarrow{SA_3} = 0.$$

Следовательно, если коэффициенты q_2, q_3 выбрать пропорциональными $\overrightarrow{SA_2}$, $\overrightarrow{SA_3}$,то координаты точек A_1 и S будут удовлетворять уравнению

$$q_2m_2 + q_3m_3 = 0.$$

Итак, мы нашли уравнение проведенной нами прямой. Если она окажется параллельной прямой A_2A_3, то в полученном уравнении следует положить, например, $q_2 = -q_3 = 1$.

Аналогичным образом можно рассуждать и в том случае, если заданная прямая проходит через другую вершину треугольника $A_1A_2A_3$. Если прямая не проходит ни через одну из вершин треугольника, то она

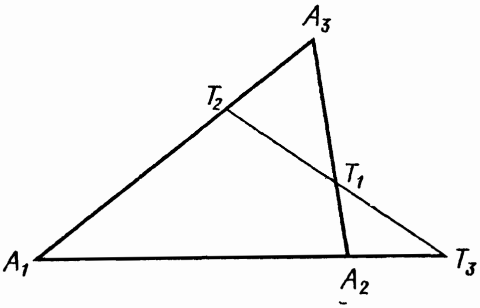


Рис. 66.

пересекает по крайней мере две его стороны, например, сторону A_2A_3 в точке T_1 и сторону A_1A_2 в точке T_2 (рис. 66).

Барицентрические координаты точки T_1 удовлетворяют соотношениям

$$m_1 = 0,\ m_2 \cdot \overrightarrow{T_1A_2} + m_3 \cdot \overrightarrow{T_1A_3} = 0,$$

а барицентрические координаты точки T_2 — соотношениям

$$m_2 = 0,\ m_3 \cdot \overrightarrow{T_2A_3} + m_1 \cdot \overrightarrow{T_2A_1} = 0.$$

Если точка T_1 должна лежать на заданной прямой

$$q_1m_1 + q_2m_2 + q_3m_3 = 0,$$

то должно выполняться равенство отношений

$$q_2 : q_3 = \overrightarrow{T_1A_2} : \overrightarrow{T_1A_3}.$$

Если точка T_2 должна лежать на прямой

$$q_1m_1 + q_2m_2 + q_3m_3 = 0,$$

то должно выполняться равенство отношений

$$q_3 : q_1 = \overrightarrow{T_2A_3} : \overrightarrow{T_2A_1}.$$

Остается лишь найти q_1, q_2, q_3, и уравнение заданной прямой будет полностью определено.

Как видно из рис. 66, должно выполняться еще одно соотношение:

$$q_1 : q_2 = \overrightarrow{T_3A_1} : \overrightarrow{T_3A_2}.$$

Перемножив отдельно левые и правые части всех трех равенств, получим

$$1 = (q_2 : q_3) \cdot (q_3 : q_1) \cdot (q_1 : q_2) =$$
$$= (\overrightarrow{T_1A_2} : \overrightarrow{T_1A_3}) \cdot (\overrightarrow{T_2A_3} : \overrightarrow{T_2A_1}) \cdot (\overrightarrow{T_3A_1} : \overrightarrow{T_3A_2}),$$

откуда

$$\overrightarrow{T_1A_2} \cdot \overrightarrow{T_2A_3} \cdot \overrightarrow{T_3A_1} = \overrightarrow{T_1A_3} \cdot \overrightarrow{T_2A_1} \cdot \overrightarrow{T_3A_2}.$$

В честь итальянского математика XVII в. последнее соотношение названо теоремой Чевы. Теорема Чевы очень похожа на теорему Менелая.

Сделаем еще одно замечание, которым воспользуемся в дальнейшем. Соотношение

$$p_1m_1 + p_2m_2 + p_3m_3 = p_0 \qquad (*)$$

также представляет собой уравнение прямой, поскольку оно эквивалентно уравнению (***). Подставив в его правую часть p_0' вместо p_0, мы получим новое уравнение

$$p_1m_1 + p_2m_2 + p_3m_3 = p_0'.$$

Если $p_0 \neq 0$, то уравнения противоречивы. В этом случае не существует системы грузов m_1, m_2, m_3, удовлетворяющих одновременно старому и новому уравнениям, и поэтому не существует и точки, лежащей одновременно на обеих прямых.

Следовательно, при $p_0 \neq p_0'$ мы получаем уравнения параллельных прямых. Итак, при различных p_0 уравнение (*) описывает пучок параллельных прямых.

Линейное программирование

Шоколад делается из какао, сахара и молока. От соотношения этих ингредиентов зависит его сорт. Каждому сорту соответствует своя рецептура, своя пропорция между составляющими.

Пусть $m_1 : m_2 : m_3$ (какао : сахар : молоко) — соотношение между весовыми долями составных частей шоколада. Начертим на плоскости треугольник $A_1A_2A_3$. Какао соответствует вершина A_1, сахару — вершина A_2, молоку — вершина A_3. Шоколаду с соотношением между весовыми долями какао, сахара и молока $m_1 : m_2 : m_3$ сопоставим центр тяжести системы грузов m_1, m_2, m_3, расставленных в вершинах A_1, A_2, A_3.

Каждая рецептура шоколада представлена одной из точек внутри треугольника $A_1A_2A_3$. Точки, лежащие вне треугольника $A_1A_2A_3$, кондитера не интересуют, поскольку у таких точек по крайней мере один из грузов m_1, m_2, m_3 отрицателен.

Чистому какао соответствует точка A_1, чистому сахару — точка A_2, чистому молоку — точка A_3. Горькие сорта шоколада располагаются ближе к точке A_1, сладкие — ближе к точке A_2, сорта шоколада с повышенным содержанием молока — в окрестности точки A_3.

Подбирать рецепт шоколада можно по вкусу. Предположим, что мы составили три сорта шоколада, соответствующие точкам B_1, B_2, B_3 внутри треугольника, а нам нравится шоколад, соответствующий точке C. Можно ли составить сорт C из смеси сортов B_1, B_2, B_3 и в каких пропорциях следует для этого смешивать сорта B_1, B_2, B_3?

Пусть n_1, n_2, n_3 — количества шоколада сортов B_1, B_2, B_3, необходимые для получения сорта C. Расположим грузы n_1, n_2, n_3 в вершинах B_1, B_2, B_3, совместим центр тяжести системы с точкой C (рис. 67).

Итак, задача сводится к следующему: требуется подобрать для точек B_1, B_2, B_3 такую систему грузов n_1, n_2, n_3, чтобы ее центр совпадал с заданной точкой C.

Разумеется, грузы n_1, n_2, n_3 не должны быть отрицательными. Но это означает, что точка C должна лежать внутри треугольника $B_1B_2B_3$.

Итак, из шоколада сортов B_1, B_2, B_3 мы сможем составить лишь такие сорта, которые соответствуют точкам, лежащим внутри треугольника $B_1B_2B_3$.

Сколько стоит шоколад? Цена шоколада зависит от того, сколько стоят его компоненты. Пусть p_1 —

стоимость единичной порции какао, p_2 — стоимость единичной порции сахара и p_3 — стоимость единичной порции молока. (В стоимость шоколада входит не только стоимость ингредиентов, но и затраты на их переработку, не зависящие от сорта шоколада. Эти затраты мы для простоты не будем учитывать в своих рассуждениях.)

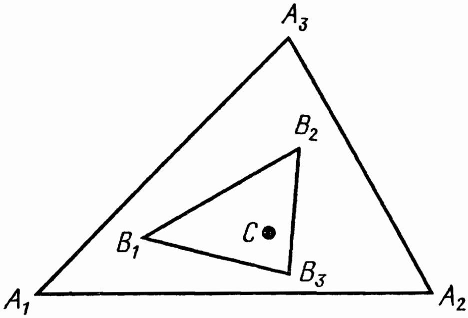


Рис. 67.

Стоимость единичной порции шоколада сорта $m_1 : m_2 : m_3$ можно представить в виде

$$p_1 m_1 + p_2 m_2 + p_3 m_3.$$

Для всех сортов шоколада, продаваемых по цене p_0, справедливо соотношение

$$p_1 m_1 + p_2 m_2 + p_3 m_3 = p_0.$$

Это означает, что точки, соответствующие сортам шоколада одинаковой стоимости, p_0 лежат на одной прямой. В конце предыдущего раздела мы упомянули о том, что прямые с различной правой частью (сорта различной стоимости) p_0 и p_0' должны быть параллельны. Предположим, что цены на какао, сахар и молоко относятся между собой, как $1:3:5$ (разумеется, наше предположение не соответствует действительности).

Такое соотношение между стоимостью единичных порций составных частей шоколада изображено на рис. 68. Прежде всего определим прямую, соответствующую $p_0 = 0$. Она должна пересекаться с прямой

A_2A_3 в такой точке T_1, для которой выполняется равенство отношений $\overrightarrow{T_1A_2} : \overrightarrow{T_1A_3} = 3:5$. Кроме того, прямая с $p_0 = 0$ должна пересекаться с прямой A_3A_1 в такой точке T_2, для которой $\overrightarrow{T_2A_3} : \overrightarrow{T_2A_1} = 5:1$. Этими двумя точками интересующая нас «прямая цен» полностью определена. Разумеется, ее точки не соответствуют реальным сортам шоколада, поскольку она целиком проходит вне треугольника $A_1A_2A_3$. Но прямые, соответствующие любым другим ценам, должны

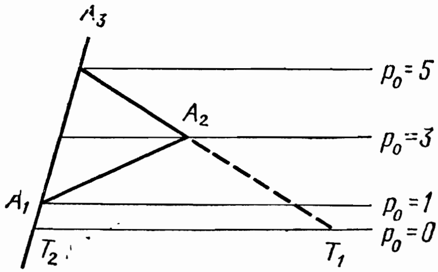


Рис. 68.

быть параллельны ей. При этом прямая с $p_0 = 1$ проходит через вершину A_1, прямая с $p_0 = 3$ — через вершину A_2, а прямая с $p_0 = 5$ — через вершину A_3. При установленном нами распределении цен самый дешевый шоколад состоит из чистого какао, а самый дорогой — из чистого молока. Если провести более густую сеть прямых, параллельных прямой с $p_0 = 0$, то можно будет узнать стоимость любого сорта шоколада.

Но оставим шоколад и обратимся к другой задаче.

Фабрика перерабатывает некоторое сырье, причем ежедневно в одном и том же количестве, которое мы примем за единицу. Переработка сырья осуществляется по трем различным технологиям, и фабрика соответственно выпускает три типа готовых изделий: I, II, III, которые приносят ей различный доход. Пусть p_1, p_2, p_3 — чистая прибыль, извлекаемая фабрикой при переработке единицы сырья по I, II и III технологиии. Если бы фабрика производила лишь изделия типа I, то чистая прибыль, извлекаемая ею из переработки единицы сырья, составляла бы ровно p_1.

Аналогично, если бы фабрика специализировалась на выпуске изделий типа II, то чистая прибыль от переработки единицы сырья составила бы p_2, а при выпуске одних лишь изделий типа III — p_3.

В день, когда количества изделий типов I, II и III представлены в продукции фабрики как $m_1:m_2:m_3$, чистая прибыль на единицу переработанного сырья составляет

$$P = p_1m_1 + p_2m_2 + p_3m_3.$$

Изобразим состав $m_1:m_2:m_3$ дневной продукции фабрики в виде центра тяжести грузов m_1, m_2, m_3, расположенных в вершинах треугольника $A_1A_2A_3$. Линиями равной прибыли будут параллельные прямые. Если $p_1:p_2:p_3 = 1:3:5$, то можно снова воспользоваться рис. 68.

Для фабрики выгоднее всего было бы производить изделия типа III, приносящие ей наибольшую прибыль, но, к сожалению, это невозможно. Прежде чем превратиться в готовое изделие, сырье должно пройти обработку на четырех машинах. В зависимости от того, на изготовление какого изделия идет сырье, обработка его на машине каждого типа требует различных затрат времени.

Продолжительность обработки сырья в зависимости от сорта изделия и типа машины представлена в следующей таблице (T, U, V, W — обозначения машин):

	I	II	III
T	t_1	t_2	t_3
U	u_1	u_2	u_3
V	v_1	v_2	v_3
W	w_1	w_2	w_3

Если требуется изготовить изделие типа I, то продолжительность обработки сырья на машине T составляет t_1, на машине U — u_1 и т. д. Если требуется изготовить изделие типа II, то машина T затратит на обработку сырья t_2 и т. д. (Время обработки сырья мы указываем в днях.) Разумеется, продолжительность работы каждой из машин T, U, V, W в течение одного дня не превышает единицы машинного времени.

Если состав дневной продукции фабрики определяется отношениями $m_1 : m_2 : m_3$, то продолжительность использования в этот день машины будет

$$t_1m_1 + t_2m_2 + t_3m_3.$$

Следовательно, должно выполняться неравенство

$$t_1m_1 + t_2m_2 + t_3m_3 \leqslant 1.$$

Аналогичные неравенства можно написать и для остальных машин:

$$u_1m_1 + u_2m_2 + u_3m_3 \leqslant 1,$$

$$v_1m_1 + v_2m_2 + v_3m_3 \leqslant 1,$$

$$w_1m_1 + w_2m_2 + w_3m_3 \leqslant 1.$$

Этими неравенствами и объясняется, почему мы не можем произвольно устанавливать состав $m_1 : m_2 : m_3$ готовой продукции: числа m_1, m_2, m_3 должны удовлетворять всем неравенствам, поскольку те учитывают производительность машин T, U, V, W.

Где же находятся внутри треугольника точки, удовлетворяющие всем неравенствам? Уравнение

$$t_1m_1 + t_2m_2 + t_3m_3 = 1$$

соответствует прямой. Точки, для которых выполняется неравенство

$$t_1m_1 + t_2m_2 + t_3m_3 < 1$$

или

$$t_1m_1 + t_2m_2 + t_3m_3 > 1,$$

расположены в одной из полуплоскостей, на которые разбивает плоскость прямая

$$t_1m_1 + t_2m_2 + t_3m_3 = 1.$$

Неравенству

$$t_1m_1 + t_2m_2 + t_3m_3 \leqslant 1$$

соответствуют не только внутренние, но и граничные точки полуплоскости.

Итак, требуется найти геометрическое место точек, барицентрические координаты которых удовлетворяют четырем неравенствам, и построить полуплоскость, соответствующую каждому из неравенств. То, что при этом получается, изображено на рис. 69. Машинам

T, U, V, W соответствуют граничные прямые T, U, V, W. Полуплоскость, удовлетворяющая соответствующему неравенству, показана штриховыми «флажками». Область внутри треугольника, принадлежащая всем четырем полуплоскостям, полностью заштрихована. Все ее точки принадлежат одновременно и треугольнику, и четырем полуплоскостям. Такая область называется *пересечением* четырех полуплоскостей и треугольника.

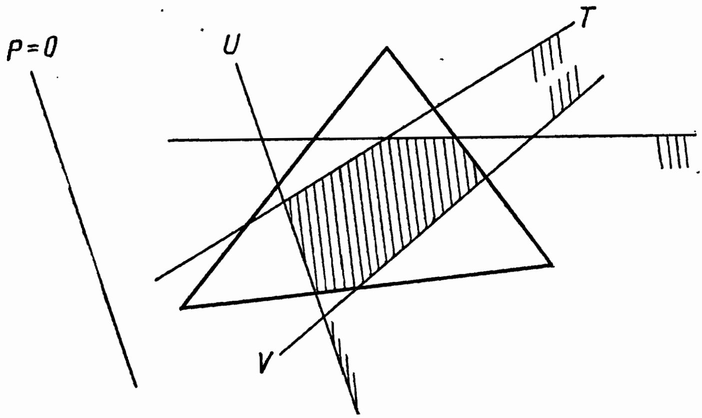


Рис. 69.

Заполняют построенную нами область точки с допустимыми значениями отношений $m_1 : m_2 : m_3$. Последнее означает, что производительность машин T, U, V, W позволяет фабрике выпускать продукцию, состав которой определен отношениями $m_1 : m_2 : m_3$. Если точка, соответствующая допустимому составу продукции, лежит, например, на прямой T, то это означает, что для выработки продукции данного состава машина T должна работать на полную мощь. На рис. 69 мы видим, что при любом допустимом ассортименте готовой продукции на полную мощность (при заданных нами условиях) могут работать не более двух машин.

Граничные прямые могли бы расположиться на плоскости так, что соответствующие им неравенства не имели бы общих решений. Например, если $t_1 > 1$, $t_2 > 1$, $t_3 > 1$, то $t_1m_1 + t_2m_2 + t_3m_3 > m_1 + m_2 + m_3 =$ $= 1$. Следовательно, затраты машинного времени были бы больше 1, в то время как они никогда не должны превышать ее. Правда, столь обескураживающий результат может означать, что мы неправильно выбрали за единицу времени день. При ином выборе единицы измерения машинного времени результат мог бы оказаться другим, хотя и не обязательно.

На рис. 69 мы видим прямую $P = 0$ (величина $P = p_1m_1 + p_2m_2 + p_3m_3$ по-прежнему означает чистую прибыль). Руководство фабрики стремится сделать ее как можно больше. Это означает, что прямую, параллельную прямой $P = 0$, следует провести через точку заштрихованной области, наиболее удаленную от прямой $P = 0$. Пусть $P = k$ — уравнение такой прямой. Любая лежащая на ней допустимая точка соответствует оптимальному (приносящему наибольшую прибыль) ассортименту продукции $m_1 : m_2 : m_3$. Такая точка может быть одна, но бывает и несколько.

Задачи, аналогичные только что рассмотренной, ныне часто встречаются в прикладной математике. Постановку и решение таких задач принято называть *программированием.* В том случае, если приходится решать системы линейных уравнений и неравенств (как это делали мы), говорят о *линейном программировании.* Рассмотренная нами задача с тремя неизвестными m_1, m_2, m_3 особенно проста. На практике приходится решать задачи, содержащие сотни неизвестных, и прибегать к услугам электронных вычислительных машин. Решение задач с большим числом неизвестных обходится недешево, но затраты на поиск оптимального решения окупаются с лихвой: затратив тысячи, мы выигрываем сотни тысяч или миллионы.

Приведем еще несколько примеров задач (без решений), встречающихся на практике.

Прежде всего нельзя не упомянуть о транспортной задаче. Компания владеет десятью предприятиями $A_1, \ldots, A_{10}$, производящими какой-то продукт в количествах $m_1, \ldots, m_{10}$. Той же компании принадлежат 50 магазинов $B_1, \ldots, B_{50}$, способных продать производимый продукт в количествах $n_1, \ldots, n_{50}$. Разумеется, во избежание затоваривания должно выполняться равенство $m_1 + \ldots + m_{10} = n_1 + \ldots + n_{50}$. Транспортные расходы известны. Стоимость перевозки единичного количества продукта с предприятия A_i в магазин B_j

составляет u_{ij}. Предположим, что с предприятия A_i в магазин B_j собираются вывести x_{ij} продукта. Как следует выбрать неизвестные для того, чтобы расходы на производство всей продукции были минимальны?

Поскольку предприятие A_i производит количество продукта m_i, то должно выполняться равенство

$$x_{i1} + \ldots + x_{i50} = m_i \quad \text{при} \quad i = 1, \ldots, 10.$$

Поскольку магазин B_j может распродать количество продукта n_j, то должно выполняться равенство

$$x_{1j} + \ldots + x_{10j} = n_j \quad \text{при} \quad j = 1, \ldots, 50.$$

Мы получаем систему из $10 + 50$ уравнений с 500 неизвестными x_{ij}. Транспортные расходы компании составляют

$$x_{11}u_{11} + \ldots + x_{10}u_{10\,1} + \ldots + x_{10\,50}u_{10\,50}.$$

Эту сумму необходимо минимизировать.

Складская задача. Для хранения определенных товаров магазин располагает особым помещением, вместимость которого ограничена. Цены на товары подвержены, как известно, сезонным колебаниям. Сбыт товаров также находится в определенной зависимости от времени года. Требуется разработать оптимальный план закупок, хранения и продажи товаров.

Задача о ресторане. Ресторан сдает грязные скатерти и салфетки в прачечную. Обычная стирка занимает неделю, срочная — два дня, но стоит дороже. Если все столовое белье отдавать в обычную стирку, то придется завести больше скатертей и салфеток, чем при срочной стирке. Сколько скатертей и салфеток следует приобрести ресторану, какую часть столового белья отдавать в обычную стирку и какую в срочную, чтобы расходы на приобретения и стирку скатертей и салфеток были минимальными?

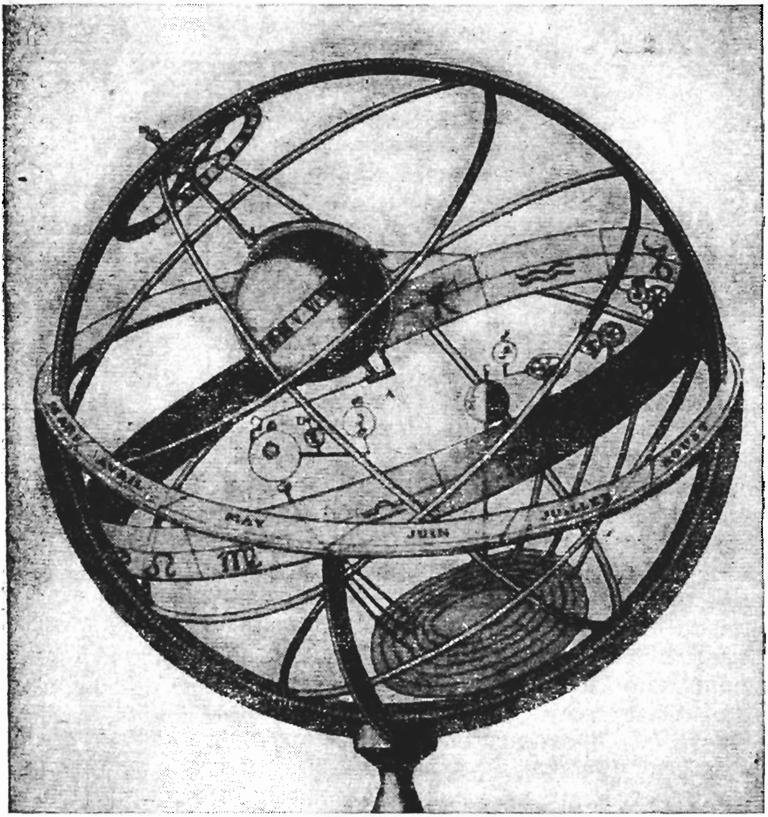
Задача коммивояжера. Коммивояжер должен за время командировки посетить 50 городов. Как ему выбрать оптимальный маршрут? Эта задада до сих пор не решена.

Шуточная задача. Можно ли считать моногамию оптимальной системой брака? Население страны состоит из мужчин A_i и женщин B_j. Взаимная симпатия мужчины A_i и женщины B_j оценивается числом u_{ij}. Если A_i проводит с B_j время t_{ij}, то сумма всех симпатий по стране выражается суммой членов вида $t_{ij}u_{ij}$. Эту сумму необходимо максимизировать. Можно показать, что максимум достигается лишь в том случае, когда каждый мужчина все свое время проводит лишь с одной женщиной, а каждая женщина — лишь с одним мужчиной. Таким образом, моногамия по праву может быть названа оптимальной системой брака.

Впрочем, моногамия может оказаться наихудшей системой брака — стоит лишь иначе составить брачные пары.

Глава 7

Мир в зеркале



Отображения в пространстве

Озеро отражает окрестные пейзажи, несколько искажая их, если разыгравшийся ветер поднимет рябь на поверхности воды. Чем глаже зеркало озера, тем точнее соответствие между оригиналом и изображением. Недаром поверхность воды называют зеркалом: реальность предстает в ней как отражение, как зеркальное отражение.

Я подхожу к зеркалу. Из его глубин навстречу мне идет незнакомец, похожий на меня, моего роста, моей комплекции, с моим лбом, моими ресницами, пальцами. Я отступаю от зеркала на шаг, и он делает назад шаг такой же длины. Я прикладываю под прямым углом к зеркалу линейку длиной 30 см. Линейка в зазеркалье, также длиной 30 см, составляет как бы продолжение настоящей. Каждый ее сантиметр находится в строгом соответствии со своим прототипом. Отражение любой точки находится в зазеркалье на таком же расстоянии от зеркала, на каком по эту сторону зеркала находится сама точка. Расстояние между любыми двумя точками перед зеркалом в точности равно расстоянию между их отражениями.

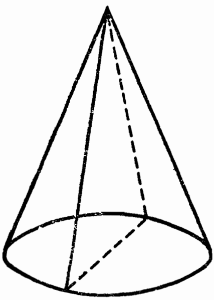


Рис. 70.

Существуют и невидимые зеркала. Одно такое зеркало я как бы всегда ношу с собой. Вертикальная плоскость, якобы проходящая через мои макушку, носовую перегородку, острие подбородка, делит пополам лицо, череп, тело, так что моя левая половина зеркально симметрична правой. Правда, при ближайшем рассмотрении симметрия оказывается неполной: мое левое ухо заострено чуть больше правого, а правый глаз видит чуть лучше левого, слева у меня расположено сердце, а справа печень. Геометрические тела обладают более совершенной симметрией. Любая плоскость, проходящая через ось кругового конуса, делит его на две зеркально-симметричные половинки (рис. 70). На рис. 71 показано, как куб можно рассечь

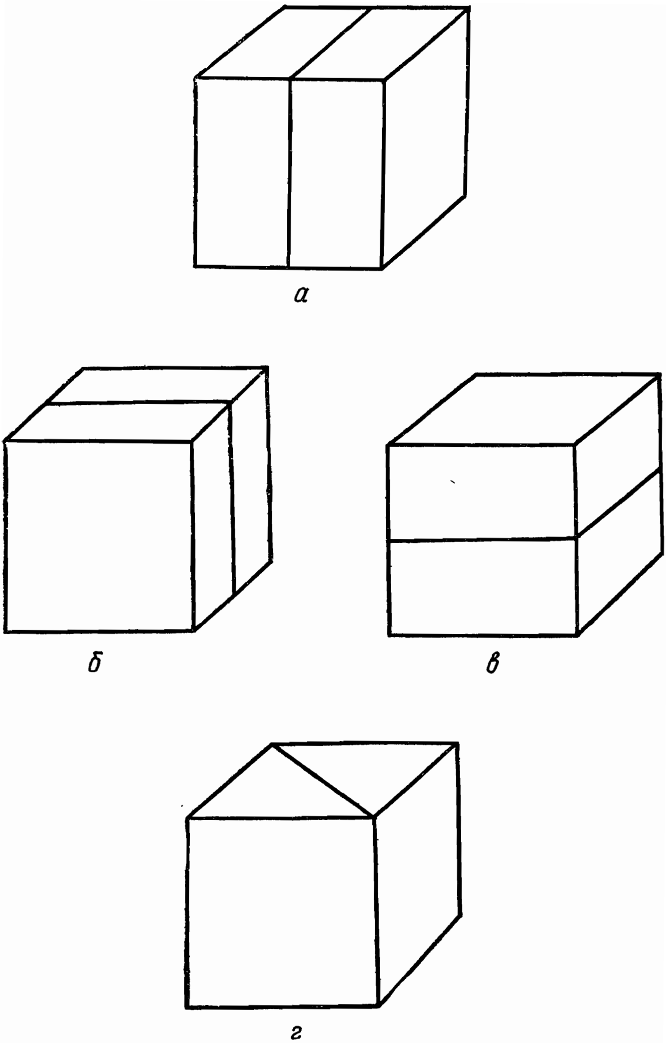


Рис. 71.

плоскостью на правую и левую (*а*), переднюю и заднюю (*б*), верхнюю и нижнюю (*в*) половинки, и всякий раз обе половинки будут зеркально-симметричны. Более того, куб можно рассечь плоскостью и на два клина (*г*), причем шестью разными способами, соответствие между ребрами неизменно будет таким, что оба клина можно совместить при отражении в зеркале.

Плоскости, рассекающие какое-нибудь тело на зеркально-симметричные половины, называются *плоскостями симметрии*. Куб обладает $3 + 6 = 9$ плоскостями симметрии.

Отражения на плоскости

Спуск на плоскость из пространства имеет определенные преимущества: рисовать плоские фигуры, например «лик Луны» (рис. 72), рассеченный пунктирной прямой на зеркально-симметричные половины, или правильный шестиугольник, который можно разрезать на зеркально-симметричные половины шестью различными способами (вдоль трех прямых, проходящих через середины противоположных сторон, — рис. 73, *а*, и вдоль трех прямых, проходящих через противоположные вершины, — рис. 73, *б*), гораздо легче, чем объемные тела. Если трехмерные тела обладали двумерными плоскостями симметрии, то двумерные фигуры обладают одномерной осью симметрии.

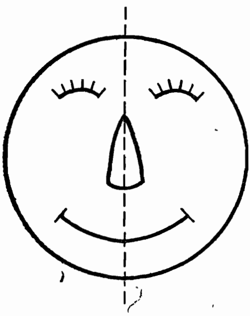


Рис. 72.

На рис. 74 изображена половина кленового листа, разрезанного по оси симметрии. Как закончить рисунок? Из точки *a* на краю листа опустим перпендикуляр на ось симметрии. Продолжив перпендикуляр за точку пересечения с осью симметрии на расстояние, равное его длине, мы получим точку *a′*, симметричную точке *a*. Проделав аналогичную операцию со всеми вершинами и соединив полученные точки, мы нарисуем контур недостававшей половины листа.

Проще всего строить точку, симметричную данной точке a относительно той или иной оси, при помощи циркуля (рис. 75). Проведем окружность с центром в

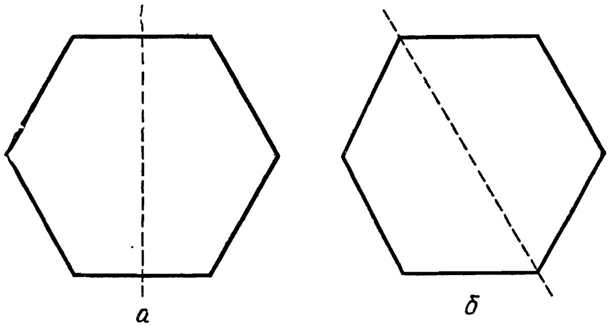


Рис. 73.

точке a, пересекающую ось симметрии S в точках b и c. Не меняя раствора циркуля, сделаем засечки из точек b и c и получим точку a'. Тем самым точка a', симметричная точке a относительно оси S, построена.

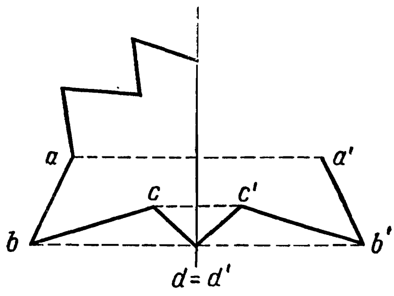


Рис. 74.

Начав построение с точки a', мы пришли бы к точке a. Следовательно, точка a симметрична точке a' относительно оси S. Итак, точки a и a' симметричны относительно оси S.

Точка P, расположенная на самой оси симметрии S, совпадает с симметричной ей точкой. Прямая A, параллельная оси S, при отражении переходит в прямую A', параллельную A и S и расположенную по другую сторону от оси симметрии S на таком же расстоянии, как прямая A (рис. 76). Если прямая A пересекает ось симметрии S в точке a, то и симметричная ей прямая A' также пересекает ось симметрии S в точке a, причем обе прямые A и A' образуют с осью симметрии S равные углы (рис. 77). Треугольник abc при отражении относительно прямой переходит в треугольник $a'b'c'$, стороны и углы которого равны соответственным сторонам и углам треугольника abc (рис. 78).

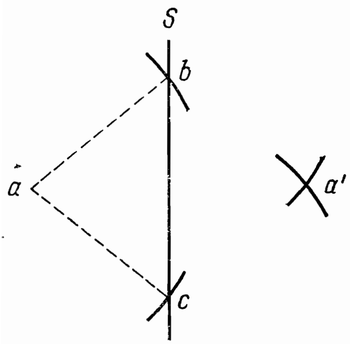


Рис. 75.

Вместо длинного выражения «точка, симметричная точке a относительно прямой S», условимся кратко записывать Sa (рис. 79). Аналогично SA означает прямую, симметричную прямой A относительно прямой S.

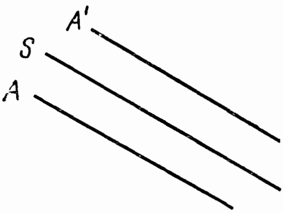


Рис. 76.

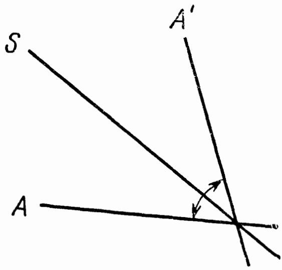


Рис. 77.

При повторном отражении относительно прямой S мы возвращаемся в исходную точку. Следовательно,

$$SSa = a,$$

$$SSA = A.$$

Обозначим расст(a, b) расстояние между точками a и b. Тогда

$$\text{расст}\ (Sa,\ Sb) = \text{расст}\ (a,\ b)$$

и аналогично

$$\angle (SA, SB) = \angle (A, B),$$

$Sa = a$, если точка a принадлежит S,
$SA = A$, если $A = S$ или $A \perp S$.

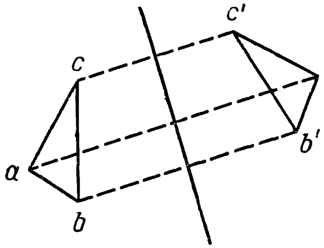


Рис. 78.

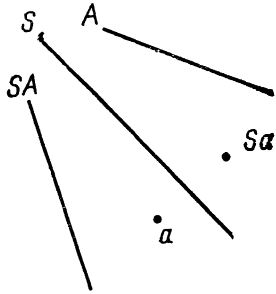


Рис. 79.

Произведение отражений относительно параллельных прямых

Рассмотрим два последовательных отражения: сначала относительно прямой S_1, а затем относительно прямой S_2. Что у нас получится?

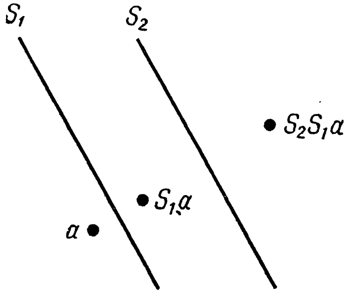


Рис. 80.

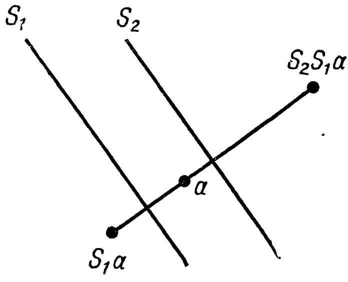


Рис. 81.

Предположим прежде всего, что прямые S_1 и S_2 параллельны и отстоят друг от друга на расстоянии, равном 1 (рис. 80). Суммарная длина перпендикуляров, опущенных из точки S_1a на прямые S_1 и S_2, равна 1. Точки a и S_1a находятся по разные стороны от прямой S_1 на равном расстоянии от нее. Точки S_2S_1a и S_1a находятся по разные стороны от прямой S_2 на равном расстоянии от нее. Следовательно, расстояние между точками a и S_2S_1a вдвое больше расстояния между прямыми S_1 и S_2, то есть равно 2. Отрезок прямой, соединяющий точки a и S_2S_1a, перпендикулярен прямым S_1 и S_2 и направлен от S_1 к S_2. Если все точки плоскости мы отразим сначала относительно прямой S_1, а затем относительно прямой S_2, то каждая точка сдвинется на расстояние, равное 2, в направлении от S_1 к S_2, перпендикулярном этим прямым.

Впрочем, мы доказали наше утверждение лишь для точек, расположенных относительно прямых S_1 и S_2 так, как точка a на рис. 84, то есть не лежащих между прямыми S_1 и S_2. Но точка a может находиться и между прямыми S_1 и S_2, как показано, например, на рис. 81, или даже лежать за прямой S_2 (по другую сторону от прямой S_1). Чтобы убедиться в правильности нашего утверждения в общем случае, проведем прямые, перпендикулярные прямым S_1 и S_2 и направленные от прямой S_1 к прямой S_2, и будем считать расстояния, измеряемые вдоль этих прямых в направлении от S_1 к S_2, положительными, а расстояния, измеряемые в противоположном направлении, — отрицательными. Отрезки, перпендикулярные прямым S_1 и S_2 и направленные от S_1 к S_2, будут иметь длину $+1$.

С такими направленными расстояниями мы и будем теперь работать. Пусть $\overrightarrow{\text{расст}}(a, b)$ — расстояние от точки a до точки b. Если три точки a, b, c лежат на одной прямой, перпендикулярной S_1 и S_2, то соотношение

$$\overrightarrow{\text{расст}}(a,\ b) + \overrightarrow{\text{расст}}(b,\ c) = \overrightarrow{\text{расст}}(a,\ c)$$

выполняется независимо от того, в какой последовательности точки a, b, c располагаются на прямой. Пусть a_1, a_2 — основания перпендикуляров, опущенных из точки a на прямые S_1 и S_2. Тогда соотношение

$$\overrightarrow{\text{расст}}(a_1,\ S_1a) + \overrightarrow{\text{расст}}(S_1a,\ a_2) = 1$$

выполняется независимо от того, где расположена точка a. Кроме того,

$$\overrightarrow{\text{расст}}(a,\ a_1) = \overrightarrow{\text{расст}}(a_1,\ S_1a),$$

$$\overrightarrow{\text{расст}}(S_1a,\ a_2) = \overrightarrow{\text{расст}}(a_2,\ S_2S_1a).$$

Наконец,

$$\overrightarrow{\text{расст}}(a,\ S_2S_1a) = \overrightarrow{\text{расст}}(a,\ a_1) + \overrightarrow{\text{расст}}(a_1,\ S_1a) +$$
$$+ \overrightarrow{\text{расст}}(S_1a,\ a_2) + \overrightarrow{\text{расст}}(a_2,\ S_2S_1a) =$$
$$= 2\,\overrightarrow{\text{расст}}(a_1,\ S_1a) + 2\,\overrightarrow{\text{расст}}(S_1a,\ a_2) = 2\cdot(+1).$$

Итак, точку, в которую переходит точка a при отражении сначала относительно прямой S_1, а затем S_2, мы можем теперь построить проще, чем раньше. Проведем из точки a перпендикулярно прямым S_1 и S_2 вектор, длина которого вдвое превышает расстояние между прямыми S_1 и S_2, и направим его от S_1 к S_2. Конец вектора совпадает с точкой S_2S_1a.

Точка S_1S_2a отлична от точки S_2S_1a, поскольку для построения точки S_1S_2a нам пришлось бы повернуть вектор в обратную сторону: от S_2 к S_1.

Если прямые S_1 и S_2 совпадают, то расстояние между ними равно 0, вектор стягивается в точку и поэтому $S_2S_1a = a$ (результат, уже известный нам из предыдущего раздела).

Сдвиги

Итак, производя одно за другим отражения относительно нескольких параллельных прямых, мы получаем в итоге не отражение, а *параллельный перенос*, называемый также *сдвигом*.

Соединив две любые точки плоскости с точками, симметричными им относительно прямой, мы получим параллельные, равные и одинаково направленные отрезки.

Таким образом, отражения относительно прямой представляют собой не что иное, как отображения, причем отображения однозначные, одного множества точек на другое множество точек той же плоскости, определения которых мы ввели в гл. 5.

Пусть Ta, Tb, ... — образы точек a, b, ... при сдвигах (рис. 82).

Все отрезки $(a,\ Ta)$, соединяющие точки с их образами при сдвигах, параллельны, равны и имеют одинаковое направление. Отсюда, в частности, следует, что отрезки $(a,\ b)$ и $(Ta,\ Tb)$ равны и параллельны.

При сдвиге любой отрезок смещается параллельно самому себе. Выполнив одно за другим отражения относительно параллельных прямых S_1 и S_2, мы получим сдвиг

$$S_2S_1a = Ta.$$

Это утверждение остается в силе и в том случае, если прямые S_2 и S_1 совпадают: вектор сдвига имеет тогда нулевую длину, сдвиг оставляет на месте любую точку плоскости. Сдвиг, возникающий при последовательном

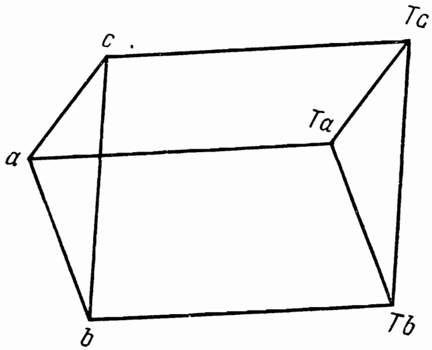


Рис. 82.

отражении относительно параллельных прямых S_1 и S_2, мы обозначим также S_2S_1 и назовем *произведением отражений* S_2, S_1 (то есть отражений относительно прямых S_2, S_1).

Произведение заданных отображений F_1 $F_2, \ldots, F_N$ снова является отображением. Обозначение его $F_NF_{N-1} \ldots F_2F_1$ надлежит понимать следующим образом. Подействовав на точку a отображением F_1, получили F_1a. Затем на F_1a подействовали F_2 и получили F_2F_1a. Так продолжали до тех пор, пока не получили $F_NF_{N-1} \ldots F_2F_1a$, подействовав F_N на $F_{N-1}F_{N-2} \ldots$ $\ldots F_2F_{1-a}$.

Сдвиг T полностью определен, если известен его вектор $\overrightarrow{(a, Ta)}$. Чтобы построить образ точки b при сдвиге T, достаточно параллельно перенести вектор $\overrightarrow{(a, Ta)}$ в точку b.

Для любого вектора всегда можно построить порождающий его сдвиг. Действительно, проведем две прямые, перпендикулярные заданному вектору и отстоящие друг от друга на расстояние, равное половине его длины. Обозначим эти прямые в том порядке, в каком мы встретим их, двигаясь по направлению, указываемому вектором, S_1 и S_2. Тогда S_2S_1 — сдвиг, порождающий исходный вектор. Заметим, что одну из прямых S_1S_2 можно выбрать произвольно (но непременно так, чтобы она была перпендикулярна вектору).

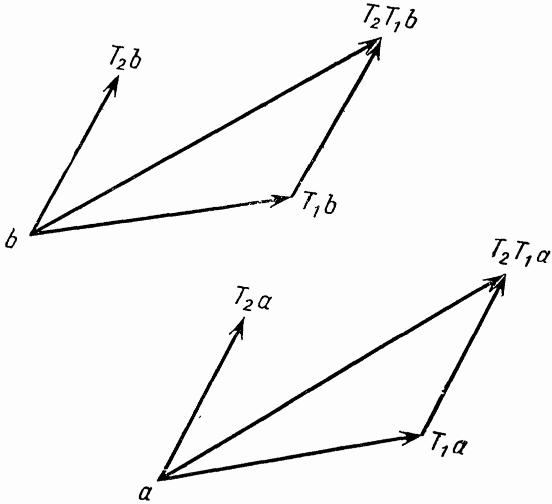


Рис. 83.

Что произойдет, если один за другим произвести два сдвига, T_1 и T_2? Сначала мы должны подвергнуть произвольно выбранную точку a сдвигу T_1 и получить точку T_1a, а затем подействовать сдвигом T_2 на T_1a. Следовательно, вектор $(\overrightarrow{a, T_2a})$ необходимо параллельно перенести в точку T_1a. Конец его совпадет с точкой T_2T_1a — образом точки a при сдвигах T_1 и T_2, выполненных один за другим. Если теперь мы захотим построить точку T_2T_1b, то сначала необходимо перенести вектор $(\overrightarrow{a, T_1a})$ в точку b, а затем вектор $(\overrightarrow{a, T_1a})$ — в точку T_1b. Результат получится таким же, как если бы мы, начав с вершины b, построили треугольник, равный треугольнику с вершинами в точках a, T_1a, T_2T_1a. Следовательно, векторы $(\overrightarrow{a, T_2T_1a})$ и $(\overrightarrow{b, T_2T_1b})$

параллельны и имеют одинаковую длину и одинаковое направление. Но тогда для всех точек x векторы $(x, \overrightarrow{T_2T_1x})$ параллельны и имеют одинаковую длину и одинаковое направление, то есть произведение двух сдвигов T_2, T_1 снова является сдвигом. Обозначив T_2T_1 произведение сдвигов T_1, T_2, получим

$$(T_2T_1)\,a = T_2(T_1a)$$

для всех точек a.

Произведение двух сдвигов мы получим, построив из конца вектора первого сдвига вектор второго сдвига. Из рис. 84 следует, что при этом несущественно,

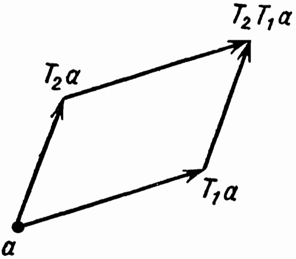


Рис. 84.

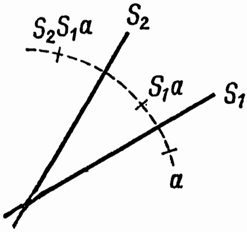


Рис. 85.

какой вектор считать первым и какой вторым: безразлично, перенесем ли мы параллельно вектор $(a, \overrightarrow{T_2a})$ в точку T_1a или вектор (a, T_1a) в точку T_2a. По построению точки T_1T_2a и T_2T_1a совпадают. Произведения сдвигов T_1T_2 и T_2T_1 являются одним и тем же сдвигом:

$$T_1T_2 = T_2T_1$$

(произведение отражений зависело от порядка сомножителей: $S_1S_2 \neq S_2S_1$).

Произведение отражений относительно пересекающихся прямых

Предположим теперь, что прямые S_1 и S_2 пересекаются (рис. 85), и произведем отражение сначала относительно прямой S_1, а затем относительно прямой S_2. Точка пересечения z при отражениях относительно обеих прямых остается неподвижной, поэтому

$$S_2S_1z = z.$$

Кроме того, и при отражении S_1 (относительно прямой S_1), и при отражении S_2 (относительно прямой S_2) все расстояния остаются неизменными, поэтому

$$\text{расст}(S_2S_1a,\ S_2S_1b) = \text{расст}(a,\ b).$$

В частности, остаются неизменными расстояния от точки z:

$$\text{расст}(S_2S_1a,\ z) = \text{расст}(a,\ z).$$

Следовательно, точки a и S_2S_1a лежат на одной и той же окружности с центром в точке z.

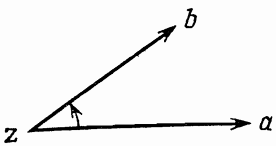


Рис. 86.

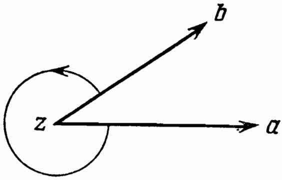


Рис. 87.

Каждая точка плоскости поворачивается вокруг точки z, поэтому все расстояния между точками плоскости остаются неизменными. Плоскость поворачивается вокруг точки z как твердое тело. Такое преобразование называется *вращением*, или *поворотом*, R вокруг точки z. Поворот возник после того, как мы сначала произвели отражение относительно прямой S_1, а затем — относительно прямой S_2, поэтому его можно представить в виде

$$R = S_2S_1,$$

то есть как произведение отражений относительно прямых S_1 и S_2.

Возникает вопрос: сколь сильно поворачивается плоскость? Как измерить поворот?

Проведем из точки z два вектора с концами в точках a и b. Мы хотим определить угол между двумя векторами, причем так, чтобы он был ориентированным.

Предположим, что первый вектор поворачивается относительно второго налево, против часовой стрелки (рис. 86). Угол, который при этом описывает первый вектор, обозначим $\angle(za, zb)$.

Следует иметь в виду, что ориентированный угол зависит от того, в каком порядке идут векторы: какой из них мы считаем первым и какой вторым. Так, если угол $\angle(za, zb)$, изображенный на рис. 86, равен α, угол $\angle(zb, za)$, как видно из рис. 87, равен $360° - \alpha$, или (поскольку в пределах одного оборота не встречаются углы, кратные 360°) углу $-\alpha$.

Если из точки z исходят три вектора, концы которых находятся в точках a, b, c, то углы между векторами всегда удовлетворяют соотношению

$$\angle(za, zb) + \angle(zb, zc) = \angle(za, zc);$$

необходимо лишь учитывать знаки углов.

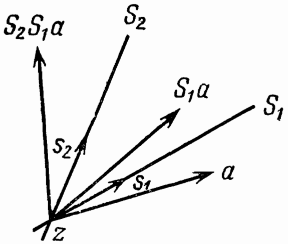


Рис. 88.

Но вернемся к отражениям относительно прямых S_1 и S_2 (рис. 88). Выберем на прямых S_1 и S_2 точки s_1 и s_2, отличные от точки z. Векторы $\overrightarrow{zs_1}$ и $\overrightarrow{zs_2}$ образуют угол φ. Кроме того, проведем из точки z векторы в точки a, S_1a и S_2a. Нас интересует угол

$$\angle(za, zS_2S_1a) = \angle(za_1, zS_1a) + \angle(zS_1a, zS_2S_1a) =$$
$$= 2\angle(zs_1, zS_1a) + \angle(zS_1a, zs_2) = 2\varphi.$$

Итак, мы доказали следующее утверждение:

при повороте $R = S_2S_1$ каждая точка движется по своей окружности с центром в точке z и поворачивается на угол, вдвое больший угла между векторами, проведенными из z вдоль прямых S_1 и S_2.

(Угол φ можно выбирать произвольно. Например, если бы мы выбрали точку s_2 на прямой S_2 по другую сторону от z, то вместо прежнего угла φ получили бы угол $180° + \varphi$. Но угол 2φ, на который поворачивается вокруг точки z любая точка плоскости, от этого не изменился бы: он оказался бы равным $2\varphi + 360°$, что совпадает с 2φ.)

Произведение отражений зависит от того, в какой последовательности мы выберем оси симметрии: сначала S_1, потом S_2 или наоборот. Выбрав сначала S_2, а затем S_1, мы заменим угол φ на угол $-\varphi$. Угол 2φ перейдет при этом в угол -2φ. Таким образом, произведение отражений S_1S_2 представляет собой поворот на угол -2φ, то есть в противоположную сторону по сравнению с поворотом S_2S_1. Углы 2φ и -2φ совпадают, лишь когда $2\varphi = 0°$ или $2\varphi = 180°$ (при $\varphi = 0°$ или $\varphi = 90°$). Следовательно,

$$S_1S_2 = S_2S_1$$

лишь в том случае, если прямые S_1, S_2 либо совпадают, либо взаимно перпендикулярны. Если прямые S_1 и S_2 совпадают, то произведение отражений относительно них порождает поворот на угол 0°, оставляющий на месте любую точку. Если прямые S_1, S_2 взаимно перпендикулярны, то произведение отражений относительно них порождает поворот на 180°, или, как еще говорят, *отражение относительно точки* z: z совпадает с серединами всех отрезков, соединяющих точки плоскости с их образами при таком повороте.

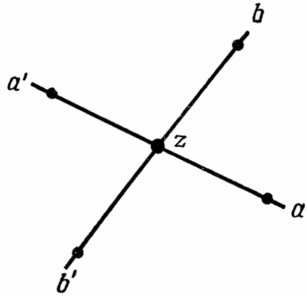


Рис. 89.

Поворот R полностью определен, если известен его *центр* z (центр окружностей, описываемых точками плоскости под действием R) и *угол поворота* φ. Любой заданный поворот R можно представить в виде произведения отражений относительно прямых S_1 и S_2, пересекающихся в его центре φ. Одну из прямых, например прямую S_1, можно выбрать произвольно (но так, чтобы она проходила через точку z), а затем позаботиться лишь о том, чтобы прямые S_2 и S_1 пересекались под углом, составляющим половину заданного угла поворота.

Произведение поворотов

Рассмотрим два поворота R_1 и R_2 вокруг центров z_1, z_2 на углы φ_1, φ_2. Мы хотим найти их произведение, то есть выполнить сначала поворот R_1, а затем поворот R_2.

Пороротт R_1 можно представить в виде произведения двух отражений:

$$R_1 = S_2 S_1.$$

Поворот R_2 также представим в виде

$$R_2 = S_4 S_3.$$

Прямой S_2 может быть любая из прямых, проходящих через точку z_1, а прямой S_3 — любая из прямых, проходящих через точку z_2. Выберем прямые S_2 и S_3 так, чтобы $S_2 = S_3$, то есть возьмем прямую, проходящую через точки z_1 и z_2, если центры поворотов R_1 и R_2 различны, и любую прямую, проходящую через точку $z_1 = z_2$, если центры поворотов R_1 и R_2 совпадают.

В произведении поворотов

$$R_2 R_1 = S_4 S_3 S_2 S_1$$

отражения S_3 и S_2, отличающиеся лишь обозначениями ($S_3 = S_2$), стоят рядом. Следовательно, мы дважды, одно за другим, производим отражение относительно одной и той же прямой, что приводит к тождественному отображению плоскости на себя. Следовательно, произведение отражений $S_3 S_2$ можно отбросить, не изменив при этом произведение поворотов $R_2 R_1$:

$$R_2 R_1 = S_4 S_1.$$

Таким образом, произведение поворотов $R_2 R_1$ представимо в виде произведения двух отражений и поэтому являет собой либо поворот, либо сдвиг. Чем именно оно является, зависит, как мы сейчас убедимся, от углов поворота φ_1 и φ_2.

Прямые S_1 и S_4 могут быть параллельными или совпадать. Тогда произведение $R_2 R_1 = S_4 S_1$ — сдвиг. Но прямые S_1 и S_4 могут и не быть параллельными. В этом случае (рис. 90), умножив R_2 на R_1, мы получим поворот вокруг точки пересечения прямых S_1 и S_4 на угол φ_3. По свойству внешнего угла треугольника $^1/_2\varphi_3 = {}^1/_2\varphi_1 + {}^1/_2\varphi_2$. Следовательно, $\varphi_3 = \varphi_1 + \varphi_2$. Равенство $\varphi_1 + \varphi_2 = 0$ возникает, лишь когда $^1/_2\varphi_1 + {}^1/_2\varphi_2 = 0$ или $^1/_2\varphi_1 + {}^1/_2\varphi_2 = 180°$. Но тогда прямые S_1 и S_4

параллельны, что соответствует случаю, когда R_2R_1 — сдвиг.

Итак, *произведение двух поворотов на угол* φ_1 *и* φ_2 *есть либо поворот на угол* $\varphi_1 + \varphi_2$ (если $\varphi_1 + \varphi_2 \neq 0$), *либо сдвиг* (*если* $\varphi_1 + \varphi_2 = 0$).

(Разумеется, может встретиться сдвиг, при котором все точки плоскости остаются на месте, но его следует рассматривать как поворот на нулевой угол.)

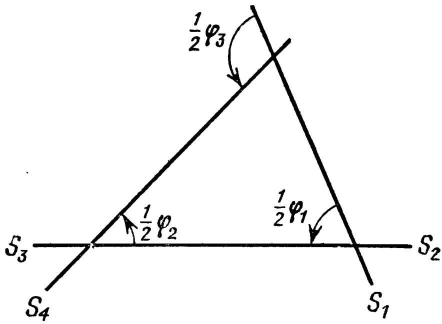


Рис. 90.

Если произведение поворотов R_2R_1 снова оказывается поворотом, то аналогичным свойством обладает и произведение поворотов R_1R_2, причем углы поворота R_2R_1 и R_1R_2 совпадают, поскольку оба угла равны сумме углов поворота R_1 и R_2, но центры поворотов R_1R_2 и R_2R_1 могут быть различными.

Если повороты R_1 и R_2 имеют общий центр, то повороты R_2R_1 и R_1R_2 также имеют общий центр (на один и тот же угол) и поэтому совпадают.

Если произведение поворотов R_2R_1 оказывается сдвигом, то произведение поворотов R_1R_2 также оказывается сдвигом (поскольку если $\varphi_1 + \varphi_2 = 0$, то $\varphi_2 + \varphi_1 = 0$), но в общем случае сдвиги R_2R_1 и R_1R_2 не совпадают.

Произведение поворота и сдвига

Пусть R — поворот, T — сдвиг. Поворот R происходит вокруг центра z на угол $\varphi \neq 0$. Обозначим S_2 прямую, проходящую через точку z перпендикулярно вектору сдвига T. Прямую S_1 можно выбрать так, чтобы

$$R = S_2S_1.$$

Сдвиг T также можно представить в виде произведения двух отражений:

$$T = S_4 S_3.$$

Ось симметрии S_3 выберем так, чтобы она совпадала с прямой S_2 (напомним, что одну из осей симметрии всегда можно выбрать произвольно). Тогда прямой S_4 может быть лишь прямая, параллельная S_3 и отстоящая от нее на расстояние, равное половине длины

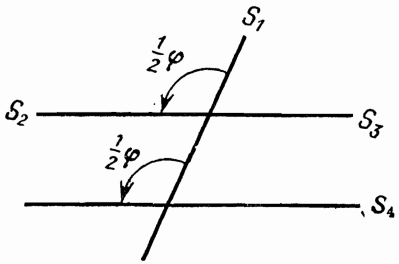


Рис. 91.

вектора сдвига (рис. 91). Это позволяет нам снова отбросить в произведении

$$TR = S_4 S_3 S_2 S_1$$

члены $S_3 S_2$, в силу чего

$$TR = S_4 S_1.$$

Поскольку прямая S_4 параллельна прямой $S_3 = S_2$, а прямые S_2 и S_1 пересекаются, то прямые S_4 и S_1 заведомо не параллельны и не совпадают. Следовательно, $S_4 S_1 = TR$ — поворот, причем на тот же угол, что и R, но вокруг другого центра.

Аналогичное утверждение справедливо и относительно произведения поворота и сдвига RT.

Итак, *произведение поворота и сдвига (или сдвига и поворота) есть снова поворот на тот же угол.*

Произведение произвольного числа отражений

Выбрав прямые S_1, S_2, S_3, S_4, можно составить произведение отражений

$$S_4 S_3 S_2 S_1.$$

Произведения отражений S_2S_1 и S_4S_3 представляют собой сдвиги или повороты (см. стр. 215—221). Следовательно, произведение двух сомножителей S_2S_1 и S_4S_3 есть либо сдвиг, либо поворот (см. стр. 222), и его также можно представить в виде произведения двух отражений.

Таким образом, произведение четырех отражений представимо в виде произведения лишь двух отражений.

Рассмотрим теперь произведение пяти отражений

$$S_5S_4S_3S_2S_1.$$

Поскольку $S_4S_3S_2S_1$ можно представить в виде произведения двух отражений, то все произведение сводится к произведению трех отражений.

Аналогично обстоит дело и с произведениями большего числа отражений. Произведение четного числа отражений всегда можно упростить и записать в виде произведения двух или нулевого числа отражений (в последнем случае мы получаем отображение, оставляющее на месте все точки плоскости). Произведение нечетного числа отражений после максимального упрощения сводится к произведению отражений, либо содержащему три сомножителя, либо вырождающемуся в один сомножитель (в последнем случае мы получаем простое отражение).

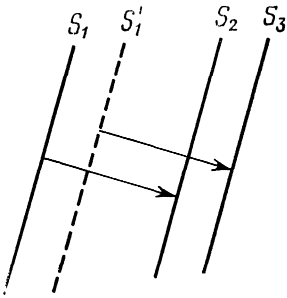


Рис. 92.

Рассмотрим произведение

$$S_3S_2S_1$$

более подробно.

Предположим сначала, что прямая S_2 параллельна прямым S_3 и S_1 (рис. 92). Тогда S_2S_1 — сдвиг. Тот же самый сдвиг мы получим, заменив прямые S_1, S_2 параллельными прямыми S_1', S_2' так, чтобы векторы, идущие от S_1 к S_2 и от S_1' к S_2', имели одинаковые длину и направление. При этом мы можем выбрать прямую S_2' так, чтобы она совпала с прямой S_3.

Тогда

$$S_3S_2S_1 = S_3S_2'S_1' = S_1'$$

— отражение.

Предположим теперь, что прямая S_2 не параллельна прямой S_3 (рис. 93). Точка пересечения z прямых S_2 и S_3 служит центром, вокруг которого происходит поворот S_3S_2. Тот же поворот мы получим, заменив

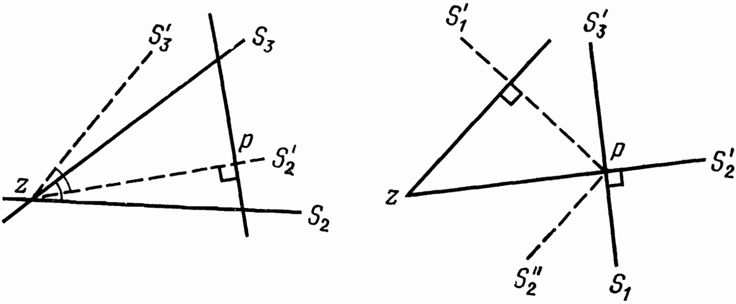


Рис. 93.

прямые S_2, S_3 прямыми S_2', S_3', пересекающимися в точке z под тем же углом, что и прямые S_2, S_3. Выберем прямую S_2' так, чтобы она была перпендикулярна прямой S_1. Рассмотрим в произведении

$$S_3S_2S_1 = S_3'S_2'S_1'$$

сомножитель $S_2'S_1$. Прямые S_1 и S_2' пересекаются в точке под прямым углом. Следовательно, $S_2'S_1$ — отражение относительно точки p. То же самое отражение относительно точки p мы получим, заменив прямые S_1, S_2' двумя другими прямыми S_1', S_2'', пересекающимися под прямым углом в точке p. Выберем прямую S_1' так, чтобы она была перпендикулярна прямой S_3'. Тогда

$$S_3S_2S_1 = S_3'S_2'S_1 = S_3'S_2''S_1,$$

а поскольку при этом прямая S_1' перпендикулярна прямым S_2'' и S_3', то прямые S_2'' и S_3' параллельны. Заметим, что

$$T = S_3'S_2''$$

— сдвиг и

$$S_3S_2S_1 = TS$$

(мы обозначили S прямую S') — произведение отражения и сдвига, вектор которого параллелен оси симметрии S (рис. 94). Такое отображение называется сдвигом с отражением. Кроме того, в рассматриваемом случае $TS = ST$. Отражение относительно прямой можно рассматривать как частный случай сдвига с отражением, соответствующий сдвигу с нулевым вектором.

Если прямая S_2 параллельна прямой S_3, то, не ограничивая общности, можно предположить, что прямая S_2 не параллельна прямой S_1 (поскольку случай,

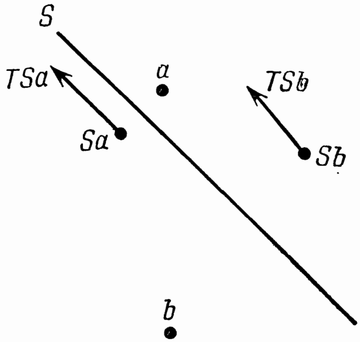


Рис. 94.

когда прямая S_2 параллельна двум другим прямым, мы уже разобрали). Проводя рассуждения, аналогичные предыдущим, мы приходим к тому же результату, что и раньше.

Итак, *произведение четного числа отражений есть либо сдвиг, либо поворот, а произведение нечетного числа отражений является скользящей симметрией.*

Отображения

Отражения, сдвиги, повороты, скользящие симметрии представляют собой не что иное, как частные случаи отображений. В общем случае, если по некоторому закону каждой точке плоскости сопоставлена некоторая другая точка плоскости, мы говорим об *отображении* плоскости на себя.

Например, следующее преобразование является отображением плоскости на себя.

Выберем некоторую точку плоскости p. Каждой точке a сопоставим точку a', лежащую на прямой pa, которая проходит через точки p и a и отстоит от p на расстояние, вдвое превышающее длину отрезка pa (рис. 95). Такое преобразование называется растяжением от точки p с коэффициентом растяжения 2. Вместо двойки можно было бы выбрать любое другое

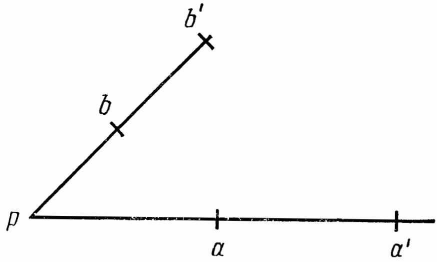


Рис. 95.

число, причем не только положительное, но и отрицательное: при отрицательном коэффициенте растяжения образ a' точки a и сама точка a будут лежать на прямой pa по разные стороны от точки p.

При растяжении в m раз все расстояния на плоскости возрастут в $|m|$ раз, где $|m|$ — абсолютная величина коэффициента растяжения m. При $|m| < 1$ отображение называют *сжатием*.

При отображениях этого рода у каждой точки плоскости существует ровно один прообраз. Следовательно, речь идет о взаимно однозначных отображениях, при которых все исходное множество отображается на все множество-образ и различным точкам последнего соответствуют различные прообразы. Взаимно однозначные отображения допускают *обращение*, то есть для каждого взаимно однозначного отображения существует некое отображение, переводящее образ любой точки снова в прообраз и притом взаимно однозначное. Если прямое отображение переводит a в a', то обратное отображение переводит a' в a.

Отображением, обратным растяжению от точки p с коэффициентом растяжения m $(m \neq 0)$, служит растяжение от той же точки p с коэффициентом растяжения $1/m$ (как было показано в главе 2, число $1/m$ обратно числу m относительно операции умножения).

Отображением, обратным сдвигу, служит сдвиг с равным по длине, но противоположно направленным вектором сдвига. Отображением, обратным повороту на угол φ вокруг центра z, служит поворот на угол $-\varphi$ вокруг того же центра z (поворот в обратную сторону). Отображение, обратное отражению относительно прямой или точки, совпадает с самим отражением. Отображением, обратным сдвигу с отражением, служит сдвиг с отражением, состоящий из обратного сдвига и отражения, совпадающего с исходным.

Если выполнить любое отображение, а затем произвести обратное отображение, то образ любой точки вернется на свое исходное место — перейдет в прообраз — и ничто не изменится. Тем не менее двойное отображение — произведение некоторого отображения и обратного ему отображения — удобно рассматривать вместе с другими отображениями. Оно называется *тождественным отображением* и соответствует растяжению с коэффициентом растяжения 1.

Поскольку коэффициентом растяжения может быть любое действительное число, то допустимо рассматривать и растяжения от точки p с нулевым коэффициентом растяжения. При таком растяжении любая точка плоскости переходит в точку p, то есть прообразом одной-единственной точки p служит вся плоскость. Нулевое отображение не взаимно однозначно. Что можно утверждать относительно его обратимости? В смысле данного нами определения нулевое отображение не имеет обратного, поскольку различные точки при таком отображении должны иметь один и тот же прообраз.

Условимся обозначать отображения заглавными латинскими буквами и, в частности, тождественное отображение буквой I. Если F — некоторое отображение и a — точка плоскости, то образ точки a при отображении F обозначим Fa. Чтобы отобразить на себя плоскость, отображение F должно быть определено для всех точек плоскости. В частности, $Ia = a$ для всех a.

Если отображение F обратимо, то обратное ему отображение обозначим F^{-1}. Отображение, обратное обратному, совпадает с исходным (прямым) отображением: $(F^{-1})^{-1} = F$ для всех a.

Пусть заданы два отображения F_1 и F_2. Подействуем на точки плоскости сначала отображением F_1, а затем отображением F_2. Тогда точка a под действием отображения F_1 перейдет в точку F_1a, а отображение F_2 переведет точку F_1a в F_2F_1a. Возникает новое отображение F_2F_1, переводящее произвольную точку плоскости a в F_2F_1a. Отображение F_2F_1 называется произведением отображений F_1 и F_2. С произведением отображений нам уже приходилось сталкиваться раньше при рассмотрении частных случаев отображений — отражений и сдвигов. Итак, по определению

$$(F_2F_1)\,a = F_2(F_1a)$$

и аналогичным образом

$$F^{-1}Fa = FF^{-1}a = Ia = a.$$

Группы

В главе 2 при рассмотрении нашей числовой системы мы ввели четыре непременных требования, предъявляемых числам, установили четыре аксиомы, без которых над числами нельзя было бы производить арифметические действия. Для операции умножения, определенной для множества всех действительных чисел, наши аксиомы сводились к следующим:

а) произведение любых двух элементов a и b принадлежит множеству;

2) выполняется ассоциативный закон умножения $(ab)c = a(bc)$;

3) существует единичный элемент e, такой, что $ae = ea = a$ для произвольного a;

4) для любого элемента a существует обратный элемент a', такой, что $a'a = aa' = e$.

Если элементы некоторого (не обязательно числового) множества удовлетворяют этим требованиям, то оно называется мультипликативной *группой* (то есть группой по умножению). Как было показано ранее, примером мультипликативной группы может служить множество действительных чисел относительно обычного умножения.

Покажем теперь, что множество обратимых отображений образует группу.

Мы уже видели, что произведение двух или большего числа сдвигов, поворотов вокруг неподвижного центра z, растяжений снова оказывается сдвигом, поворотом вокруг центра z и растяжением. Ассоциативный закон был доказан нами для отображений общего вида. За единичный элемент во множестве обратимых отображений можно принять тождественное отображение I.

Обратное отображение также принадлежит множеству обратимых отображений, если к этому множеству принадлежит прямое отображение: отображение, обратное сдвигу, — сдвиг, отображение, обратное повороту центра z, — поворот вокруг центра z в обратную сторону. Таким образом, известные нам отображения — растяжения, сдвиги и повороты вокруг общего центра z — образуют группу.

Рассмотрим теперь всевозможные произведения отражений (произведение, содержащее 0 сомножителей, по определению совпадает с тождественным отображением I). Они также образуют группу. Произведение любых двух произведений отражений само является произведением отражений. Отображением, обратным произведению отражений $S_nS_{n-1} \ldots S_2S_1$, служит произведение отражений $S_1S_2 \ldots S_{n-1}S_n$, поскольку

$$\begin{aligned} S_1S_2 \ldots S_{n-1}S_nS_nS_{n-1} \ldots S_2S_1 &= \\ = S_1S_2 \ldots S_{n-1}S_{n-1} \ldots S_2S_1 &= \\ = S_1S_2 \ldots S_2S_1 &= \\ \ldots\ldots\ldots\ldots\ldots\ldots & \\ = S_1S_2S_2S_1 &= \\ = S_1S_1 = I. & \end{aligned}$$

Таким образом, отображение, обратное произведению отражений, есть снова произведение отражений.

Отдельные отражения сами по себе не образуют группу, поскольку произведение двух отражений не является отражением.

В общем случае под *группой* отображений плоскости понимают систему Σ отображений, содержащую вместе с любыми двумя отображениями F_1, F_2 их произведение F_2F_1, тождественное отображение I и обратное отображение F^{-1} для любого принадлежащего Σ отображения F.

Если подсистема Σ' группы Σ сама является группой, то называется *подгруппой* группы Σ.

В группе, состоящей из произведений отражений, сдвиги образуют подгруппу. Произведения отражений, содержащие четное число сомножителей, также образуют подгруппу.

Автоморфизмы

Отображение плоскости, сохраняющее неизменным расстояние между любой парой точек, называется *автоморфизмом.* Под действием автоморфизма плоскость ведет себя, если можно сказать, «подобно твердому телу»: любая фигура переходит в конгруэнтную фигуру, прямая — в прямую, окружность — в окружность. Если F — автоморфизм, то для любой пары точек a, b плоскости

$$\text{расст}(Fa,\ Fb) = \text{расст}(a,\ b).$$

Автоморфизмы образуют группу, поскольку если

$$\text{расст}(F_1a,\ F_1b) = \text{расст}(a,\ b),$$
$$\text{расст}(F_2a,\ F_2b) = \text{расст}(a,\ b),$$

то

$$\text{расст}(F_2F_1a,\ F_2F_1b) = \text{расст}(F_1a,\ F_1b) = \text{расст}(a,\ b)$$

и

$$\text{расст}\left(F_1^{-1}a,\ F_1^{-1}b\right) = \text{расст}\left(F_1F_1^{-1}a,\ F_1F_1^{-1}b\right) = \text{расст}(a,\ b).$$

Пусть a, b — две различные точки плоскости и F — автоморфизм, оставляющий их неподвижными, то есть такой, что

$$Fa = a, \quad Fb = b.$$

Тогда для любой точки x (рис. 96)

$$\text{расст}(Fx,\ a) = \text{расст}(Fx,\ Fa) = \text{расст}(x,\ a),$$
$$\text{расст}(Fx,\ b) = \text{расст}(Fx,\ Fb) = \text{расст}(x,\ b).$$

Это означает, что образ Fx точки x находится от точек a и b на таких же расстояниях, как и сама точка x. Но на плоскости существует не более двух точек, находящихся на заданных расстояниях от точек a, b, причем обе эти точки симметричны относительно прямой, проходящей через точки a и b. Следовательно, если автоморфизм F оставляет неподвижными две различные точки a и b, то для каждой точки x плоскости ее образ Fx либо совпадает с ней самой, либо симметричен ей относительно прямой, проходящей через точки a и b.

Автоморфизм F может оставлять неподвижными все точки плоскости, тогда

$$F = I.$$

В противном случае существует такая точка c, которая не совпадает с Fc. Пусть x — неподвижная точка автоморфизма $F: Fx = x$. Тогда

$$\text{расст}(x, c) = \text{расст}(Fx, Fc) = \text{расст}(x, Fc).$$

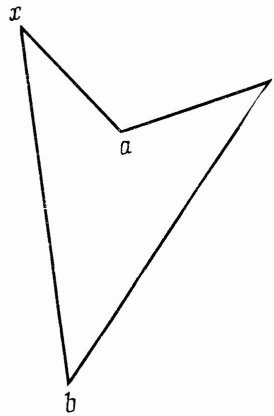


Рис. 96.

Это означает, что точка x лежит на перпендикуляре, проходящем через середину отрезка (c, Fc), то есть на прямой ab. Следовательно, все точки, не лежащие на прямой ab, не могут быть неподвижными точками автоморфизма F. Для таких точек Fx не совпадает с x, и поэтому сами точки x и их образы Fx при автоморфизме F симметричны относительно прямой ab. Для точек, расположенных на прямой ab, последнее всегда верно.

Таким образом, в рассматриваемом случае автоморфизм F оказывается не чем иным, как отражением относительно прямой ab.

Итак, *автоморфизм плоскости с двумя различными неподвижными точками a, b является либо тождественным отображением, либо отражением относительно прямой ab.*

Рассмотрим теперь две пары точек a, b и a', b', такие, что

$$\text{расст}(a,\ b) = \text{расст}(a',\ b').$$

Докажем следующее утверждение: существуют ровно два автоморфизма F, переводящих a, b в a' b', то есть таких, что

$$Fa = a', \quad Fb = b'.$$

Если $a = a'$ и $b = b'$, то тождественное отображение I — тот самый автоморфизм F, который требуется найти. Если $a \neq a'$ и $b \neq b'$ (рис. 97), то точка a

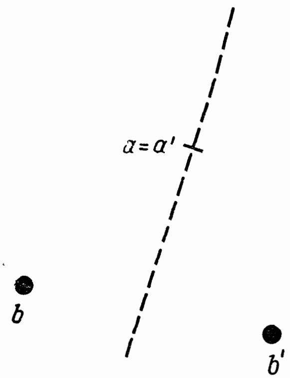


Рис. 97.

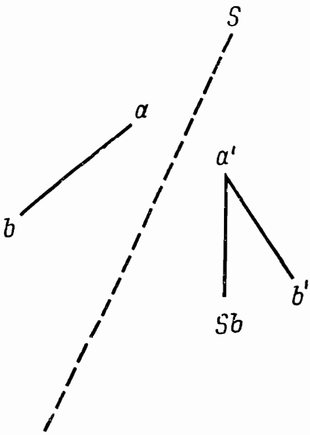


Рис. 98.

лежит на перпендикуляре, проходящем через середину отрезка bb'; и интересующим нас автоморфизмом служит отражение относительно этого перпендикуляра.

Предположим, что $a \neq a'$ и $b \neq b'$ (рис. 98). Прежде всего произведем отражение относительно перпендикуляра S, проходящего через середину отрезка aa'. Тогда

$$a' = Sa$$

и

$$\text{расст}(a',\ b') = \text{расст}(a,\ b) = \text{расст}(Sa,\ Sb) = \\ = \text{расст}(a',\ Sb).$$

Расположение пар точек a', Sb и a', b' удовлетворяет предположениям предыдущего случая. Следовательно, существует автоморфизм F_0, переводящий пару точек a', Sb в пару точек a', b':

$$F_0 a = a', \quad F_0 S b = b'.$$

Поскольку $a' = Sa$, то

$$F_0 S a = a', \quad F_0 S b = b'.$$

Таким образом, автоморфизм $F = F_0 S$ отображает a', b' и a, b. Один из автоморфизмов, переводящих одну пару точек в другую, найден.

Еще один автоморфизм мы найдем, введя отражение относительно прямой S', проходящей через точки a', b'. Автоморфизм $S'F$ также переводит a, b в a', b'. Оба автоморфизма F и $S'F$ являются произведениями отражений.

Найденными нами автоморфизмами исчерпываются все автоморфизмы, переводящие a, b в a' b'. Действительно, пусть F и F_1 — два автоморфизма, переводящих a, b в a', b':

$$Fa = F_1 a = a', \quad Fb = F_1 b = b'.$$

Тогда F^{-1} — также автоморфизм, причем

$$F^{-1} a' = a, \quad F^{-1} b' = b$$

и, следовательно,

$$F_1 F^{-1} a' = a', \quad F_1 F^{-1} b' = b'.$$

Это означает, что автоморфизм $F_1 F^{-1}$ оставляет неподвижными две различные точки a', b'. По доказанному ранее, автоморфизм $F_1 F^{-1}$ может быть либо тождественным отображением, либо отражением относительно прямой S'.

Если $F_1 F^{-1} = I$, то $F_1 F^{-1} F = IF$, $F_1 = F$.

Если же $F_1 F^{-1} = S'$, то $F_1 F^{-1} F = S'F$, $F_1 = S'F$.

Итак, мы доказали утверждение:

если a, b и a', b' — две пары точек, такие, что

$$\text{расст}(a, b) = \text{расст}(a', b') \neq 0,$$

то существуют ровно два автоморфизма, переводящие a, b в a', b'; оба автоморфизма представимы в виде произведения отражений и отличаются на одно отражение.

Отсюда следует также, что *любой автоморфизм можно представить в виде произведения отражений.*

Рассмотрим теперь три точки, a, b, c, не лежащие на одной прямой, и другие три точки, a', b', c', такие, что

$$\text{расст}(a,\ b) = \text{расст}(a',\ b'),$$
$$\text{расст}(b,\ c) = \text{расст}(b',\ c'),$$
$$\text{расст}(c,\ a) = \text{расст}(c',\ a').$$

Покажем, что существует ровно один автоморфизм, переводящий a, b, c в a', b', c'.

Как мы только что доказали, заведомо существует автоморфизм F_0, переводящий a, b в a', b':

$$F_0a = a', \quad F_0b = b'.$$

Поскольку

$$\text{расст}(F_0,\ a') = \text{расст}(F_0,\ c,\ F_0a) =$$
$$= \text{расст}(c,\ a) = \text{расст}(c',\ a'),$$

то

$$\text{расст}(F_0c,\ a') = \text{расст}(c',\ a')$$

и

$$\text{расст}(F_0c,\ b') = \text{расст}(c',\ b').$$

Таким образом, F_0c находится от a', b' на таких же расстояниях, как и точка c. Следовательно, либо $F_0c = c'$, либо точки F_0c и c' симметричны относительно прямой $a'b'$. В первом случае автоморфизм F, который переводит одну тройку точек a, b, c в другую тройку точек a', b', c', совпадает с автоморфизмом F_0, во втором случае $F = SF_0$, где S — отражение относительно прямой $a'b'$.

Если F и F_1 — автоморфизмы, переводящие a, b, c в a', b', c', то автоморфизм $F^{-1}F_1$ оставляет неподвижными точки a, b, c и совпадает либо с тождественным автоморфизмом I, либо с отражением относительно прямой ab. Вторая альтернатива отпадает, поскольку c — неподвижная точка автоморфизма $F^{-1}F_1$ и в то же время не принадлежит прямой ab. Следовательно, $F^{-1}F_1 = I$, откуда $F_1 = F$.

Итак, *существует ровно один автоморфизм, переводящий не лежащие на одной прямой точки* a, b, c в точки a', b', c', такие, что

$$\text{расст}(a,\ b) = \text{расст}(a',\ b'),$$
$$\text{расст}(b,\ c) = \text{расст}(b',\ c'),$$
$$\text{расст}(c,\ a) = \text{расст}(c',\ a').$$

Окончательный вывод: *автоморфизмы плоскости образуют группу. Ту же группу образуют произведения отражений. Любой автоморфизм допускает представление в виде произведения* 0, 1, 2 *или* 3 *отражений. Если точки* a, b, c *не лежат на одной прямой и попарные расстояния между ними совпадают с попарными расстояниями между точками* a', b', c', *то существуют ровно два автоморфизма, переводящих* a, b *в* a', b', *ровно один автоморфизм, переводящий* a, b, c *в* a', b', c'.

Движения

Рассмотрим более подробно автоморфизмы плоскости, представимые в виде произведения четного числа (0 или 2) отражений. Такие автоморфизмы называются *движениями.* Как мы успели убедиться, каждое движение — это либо сдвиг, либо поворот. Движения образуют группу, которая содержится в группе автоморфизмов (иначе говоря, движения образуют подгруппу группы автоморфизмов).

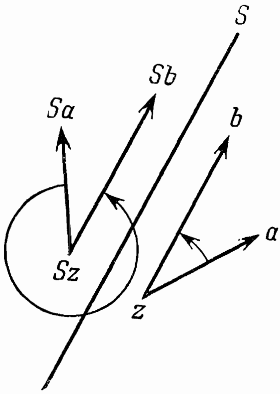


Рис. 99.

Можно ли, «взглянув» на автоморфизм, сразу, то есть не разлагая его в произведение отражений, сказать, является ли он движением?

Расстояния между любыми парами точек при отражении не изменяются, как в обычном смысле не изменяются и углы. На стр. 220 мы ввели более точное определение угла. При отражении S угол (za, zb) переходит в угол $\angle(SzSa,\ SzSb)$ (рис. 99). Угол-образ равен по величине углу-прообразу, но имеет противоположный знак. После двух отражений угол-образ имеет ту же ориентацию, что и исходный угол, после трех отражений его ориентация вновь изменяется на противоположную. Итак, автоморфизмы плоскости сохраняют расстояния. *Под действием автоморфизмов все углы сохраняют свою величину, но,* *быть может, изменяют знак. Движения отличаются от остальных автоморфизмов тем, что сохраняют не только величину, но и знак углов.*

Сохранение ориентированных углов позволяет говорить именно о движениях. Сдвиги и повороты можно производить не сразу, а постепенно, отражения и скользящие симметрии не терпят подобной нерешительности: ориентацию угла нельзя изменить на половину или три четверти, она изменяется на противоположную только целиком и сразу.

В пространстве

В пространстве кое-что происходит так же, как на плоскости, а кое-что становится более сложным. Рассмотрим кратко лишь то, что пространство в неизменном виде наследует у плоскости.

Начнем с отражений тел относительно плоскостей, которые мы будем обозначать S, S_1 и т. д.

Если плоскости S_1, S_2 параллельны, то произведение S_2S_1 отражений совпадает со сдвигом, только не плоскости, а пространства: векторы, идущие от точки x к ее образу S_2S_1x, для всех точек пространства параллельны и имеют одинаковую длину и одинаковое направление. Наоборот, любой сдвиг пространства можно представить в виде произведения отражений относительно двух надлежащим образом подобранных параллельных плоскостей. Сдвиги пространства образуют группу, и для любых двух сдвигов T_1, T_2 произведение не зависит от порядка сомножителей: $T_1T_2 = T_2T_1$.

Если плоскости S_1, S_2 не параллельны, то произведение отражений S_2S_1 совпадает с поворотом вокруг линии пересечения плоскостей S_1 и S_2 на удвоенный угол между этими плоскостями. Произведение отражений относительно плоскостей S_1, S_2 не зависит от порядка сомножителей ($S_1S_2 = S_2S_1$) тогда и только тогда, когда либо плоскости S_1 и S_2 совпадают, либо плоскость S_1 перпендикулярна плоскости S_2. В первом случае произведение двух отражений совпадает с тождественным отображением I, во втором — с отражением относительно линии пересечения плоскостей S_1 и S_2.

Отображения пространства на себя, сохраняющие расстояния между любыми парами точек, называются *автоморфизмами*. Автоморфизмы пространства образуют группу. Примерами автоморфизмов могут служить отражения относительно плоскостей и произведения отражений, образующие группу. Она входит в группу всех автоморфизмов пространства как несобственная (то есть совпадающая со всей группой) подгруппа.

Итак, *автоморфизмы пространства образуют группу. Ту же группу образуют произведения отражений. Любой автоморфизм допускает представление в виде произведения не более чем четырех отражений. Если точки a, b, c, d не лежат в одной плоскости и попарные расстояния между ними совпадают с попарными расстояниями между точками a', b', c', d', то существуют ровно два автоморфизма, переводящие a, b, c в a', b', c' (они отличаются одним множителем-отражением), и ровно один автоморфизм, переводящий a, b, c, d в a', b', c', d'!*

Доказывается эта теорема так же, как аналогичная ей теорема для плоскости, которую мы только что рассмотрели. Не более чем четырьмя последовательными отражениями относительно надлежаще выбранных плоскостей точки a, b, c, d переводят в точки a', b', c', d'. После того как точки a, b, c совмещены с точками a', b', c', отобразить точку d в d' можно двумя способами.

Как определить, что такое движения в пространстве? Произведения четного числа отражений относительно плоскостей заведомо образуют группу. Но является ли она собственной подгруппой группы автоморфизмов пространства? Иначе говоря, можно ли быть уверенным в том, что подгруппа произведений четного числа отражений не совпадает со всей группой автоморфизмов? Любой автоморфизм допускает множество разнообразнейших представлений в виде произведения отражений, и вполне можно было бы думать, что автоморфизм, представимый в виде произведения нечетного числа отражений, представим и в виде произведения четного числа отражений. На плоскости нам удалось исключить такую возможность: при четном числе отражений ориентированные углы сохранялись, при нечетном — изменяли ориентацию на противоположную. Тем самым оба типа произведений (содержащих четное и нечетное число отражений) оказались на плоскости полностью разделенными. К сожалению, угол между двумя плоскостями не обладает свойствами ориентированного угла между двумя лучами: его ориентацию изменяет отражение относительно плоскости, делящей пополам угол между плоскостями, при отражении же относительно плоскости, перпендикулярной линии пересечения образующих его плоскостей, он остается неизменным.

Попытаемся подойти к решению интересующей нас проблемы с другой стороны.

Начнем с предварительного замечания. Если S_1, S_2, S_3 — параллельные плоскости, то существует параллельная им плоскость S_4, такая, что

$$S_4S_3 = S_2S_1.$$

Необходимо позаботиться лишь о том, чтобы плоскости S_3, S_4 отстояли друг от друга на таком же расстоянии, как и плоскости S_1, S_2, и следовали в том же порядке. Еще одно аналогичное замечание. Если три плоскости S_1, S_2, S_3 пересекаются по одной прямой, то существует плоскость S_4, содержащая эту прямую и такая, что

$$S_4S_3 = S_2S_1.$$

Необходимо лишь позаботиться о том, чтобы

$$\angle (S_4, S_3) = \angle (S_2, S_1).$$

Объединим теперь оба рассмотренных нами случая. Пучком плоскостей называется либо совокупность всех плоскостей, параллельных заданной плоскости, либо совокупность всех плоскостей, проходящих через заданную прямую. Справедливо следующее утверждение: если плоскости S_1, S_2, S_3 принадлежат одному пучку, то в том же пучке найдется плоскость S_4, такая, что

$$S_4S_3 = S_2S_1.$$

Кроме того, если плоскости S_1, S_2 взаимно перпендикулярны, то

$$S_1S_2 = S_2S_1.$$

Предположим, что автоморфизм F можно представить в виде произведения отражений:

$$F = S_m S_{m-1} \dots S_2 S_1.$$

Мы будем изменять представление автоморфизма F, сохраняя число сомножителей в произведении (сам автоморфизм F также остается неизменным).

Возьмем произвольную точку пространства a. В пучке плоскостей, содержащем плоскости S_1, S_2, найдется плоскость S_1', которой принадлежит точка a. Определим плоскость S_2' так, чтобы

$$S_2' S_1' = S_2 S_1,$$

и подставим $S_2' S_1'$ в произведение $S_m S_{m-1} \dots S_2 S_1$ вместо $S_2 S_1$. Автоморфизм F тем самым мы преобразуем к виду

$$F = S_m S_{m-1} \dots S_2' S_1'.$$

Аналогично поступим с пучком плоскостей, содержащим S_2', S_3. Найдем в этом пучке плоскость S_2'', проходящую через точку a, и определим плоскость S_3' так, чтобы

$$S_3' S_2'' = S_3 S_2'.$$

После подстановки этого равенства в произведение автоморфизм F примет вид

$$F = S_m S_{m-1} \dots S_3' S_2'' S_1'.$$

Продолжая действовать в том же духе, мы в конце концов придем к следующему представлению автоморфизма F:

$$F = S_m' S_{m-1}'' \dots S_3'' S_2'' S_1', \tag{1}$$

где все плоскости, кроме, быть может, S_m', содержат точку a.

Повторив те же построения для точки b, отличной от точки a, мы получим представление автоморфизма F, в котором все плоскости, кроме, быть может, последней, содержат точку a и сверх того все плоскости, кроме, быть может, двух последних, содержат точку b. Повторим все построения еще раз для точки c (не лежащей на прямой ab), получим представление автоморфизма F в виде

$$F = S_m^* S_{m-1}^* \dots S_2^* S_1^*, \tag{2}$$

где плоскости $S_1^*, \ldots, S_{m-1}^*$ содержат точку a, плоскости $S_1^*, \ldots, S_{m-2}^*$ — точку b, плоскости $S_1^*, \ldots, S_{m-}^*$ — точку c. Все плоскости $S_1^*, \ldots, S_{m-3}^*$ содержат три не лежащие на одной прямой точки a, b, c и поэтому сливаются в одну плоскость. Таким образом, произведение отражений $S_{m-3}^* \ldots S_1^*$ совпадает либо с тождественным отображением I, либо с отражением S_1^*, откуда еще раз следует, что любой автоморфизм можно представить в виде произведения не более четырех отражений.

Выберем, например, автоморфизм $F = I$. В представлении (1) все плоскости, кроме, быть может, S'_m, содержат точку a:

$$a = Ia = Fa = S'_m S'_{m-1} \ldots S''_2 S'_1 a = S'_m a.$$

Значит, плоскость S'_m также содержит точку a.

Заменяя шаг за шагом Sb на b для плоскостей, содержащих точку b, получаем представление тождественного автоморфизма, в котором все плоскости содержат точку a и все плоскости, кроме, быть может, последней, содержат также точку b. По аналогии с приведенным выше рассуждением из того, что b — неподвижная точка автоморфизма $F: b = Ib = Fb$, следует, что и последняя плоскость содержат точку b. Наконец, повторяя доказательство для точки c, получаем представление типа (2), в котором все плоскости $S_1^*, \ldots, S_m^*$ содержат три не лежащие на одной прямой точки a, b, c и поэтому совпадают. Таким образом,

$$I = S_1^* \ldots S_m^* \quad (m \text{ сомножителей}),$$

откуда следует, что число m должно быть четным.

Итак, любое представление тождественного автоморфизма I в виде произведения отражений содержит четное число отражений.

Предположим теперь, что автоморфизм F можно представить в виде произведения отражений двумя различными способами:

$$F = S_k \ldots S_1 = S'_l \ldots S'_1.$$

Тогда

$$I = F^{-1}F = S_1 \ldots S_k S'_l \ldots S'_1.$$

Следовательно, сумма $k+l$ четна, и слагаемые k и l должны быть либо оба четными, либо оба нечетными.

Итак, *представим ли данный автоморфизм в виде произведения четного или нечетного числа отражений, не зависит от выбора его представления.*

Назовем пространственными *движениями* произведения *четного числа* отражений.

Движения образуют группу. Любое движение можно представить в виде произведения четного числа отражений (в наиболее компактной записи число отражений равно 0, 2 или 4). Произведения нечетного числа отражений движениями не является.

Исследуем теперь автоморфизм более подробно. Начнем с автоморфизмов, оставляющих неподвижной заданную точку a:

$$Fa = a,$$
$$F = S_m \ldots S_1.$$

Повторяя для точки a уже известные нам подстановки, получаем для автоморфизмов F представление (1), в котором все плоскости, кроме, быть может, S'_m, содержат точку a. Но $Fa = a$, поэтому последняя плоскость S'_m также содержит точку a.

Те же рассуждения, проведенные для точек b и c (точки a, b, c не лежат на одной прямой), позволяют представить автоморфизмы F в виде произведения отражений (2), где плоскости $S_1^*, \ldots, S_m^*$ содержат точку a, плоскости $S_1^*, \ldots, S_{m-1}^*$ — точку b и плоскости $S_1^*, \ldots, S_{m-2}^*$ — точку c. Следовательно, плоскости $S_1^*, \ldots, S_{m-2}^*$ сливаются в одну плоскость. В зависимости от того, четно или нечетно число m, произведение отражений $S_1^*, \ldots, S_m^*$ совпадает либо с тождественным отображением I, либо с отражением S_1^*. Тем самым представление (2) автоморфизмов F, оставляющих неподвижной заданную точку a, вырождается в произведение трех отражений при нечетном m и двух отражений при четном m.

Итак, *пусть F — автоморфизм с неподвижной точкой a. Тогда F допускает представление в виде произведения двух или трех отражений относительно плоскостей, содержащих точку a.*

Следовательно, *движения, оставляющие неподвижной точку a*, представимы в виде произведения двух отражений и, таким образом, совпадают с поворотами (быть может, на нулевой угол, то есть с тождественным отображением).

Рассмотрим произвольный автоморфизм пространства F, не совпадающий с чистым сдвигом или чистым поворотом. Пусть a — произвольная точка пространства. Существует сдвиг T, переводящий точку a в Fa:

$$Ta = Fa,$$

$$T^{-1}Fa = a.$$

Поскольку автоморфизм $T^{-1}F$ оставляет неподвижной точку a, то из доказанного ранее следует, что $T^{-1}F$ совпадает с некоторым поворотом R, откуда

$$F = TR,$$

то есть автоморфизм F представим в виде произведения сдвига и поворота.

Разложим сдвиг T на две составляющие, одна из которых параллельна, а другая перпендикулярна оси поворота. Пусть T_1 — сдвиг вдоль оси поворота, T_2 — сдвиг вдоль перпендикулярного направления, тогда

$$T = T_1 T_2$$

и

$$F = T_1 T_2 R.$$

Рассмотрим действие автоморфизма T_2R лишь на какой-нибудь плоскости E, перпендикулярной оси поворота R. Сдвиг T_2 и поворот R отражают плоскость E на себя, поэтому их произведение T_2R также отображает плоскость E на себя. На плоскости E автоморфизм T_2R действует как движение. Поскольку исходный автоморфизм F не совпадает с чистым сдвигом, то поворот R отличен от тождественного отображения. Следовательно, движение T_2R на плоскости E действует как поворот и имеет одну неподвижную точку b. Но тогда T_2R — автоморфизм пространства, оставляющий неподвижной точку b, то есть совпадающий с поворотом R_1. Точка b и плоскость E под действием поворота R_1 переходят в себя. Таким образом, ось поворота R_1 перпендикулярна плоскости E. Итак,

$$F = T_1 R_1,$$

где T_1 — сдвиг вдоль оси поворота R_1. Автоморфизмы, представимые в виде произведения поворота и сдвига вдоль его оси, называются *винтовым движением*. Сдвиги и повороты можно рассматривать как частные случаи винтового движения: сдвиги — при угле поворота, равном 0°, повороты — при нулевом векторе сдвига.

Итак, *все движения в пространстве — винтовые движения.*

Обратное утверждение (все винтовые движения — движения) очевидно.

Наконец, рассмотрим автоморфизмы, не являющиеся движениями. Одна разновидность таких автоморфизмов нам уже известна: это — отражения. Что можно сказать об остальных автоморфизмах — «не движениях», представимых в виде

$$F = S_3S_2S_1?$$

Если плоскости S_1, S_2, S_3 параллельны, то существует плоскость S_4, такая, что

$$S_3S_2 = S_4S_1,$$

поэтому автоморфизм

$$F = S_4S_1S_1 = S_4$$

совпадает с отражением относительно плоскости S_4. Предположим, что не все плоскости S_1, S_2, S_3 параллельны. Пусть, например, плоскости S_1 и S_2 пересекаются (остальные случаи аналогичны). Тогда существует плоскость S_2', проходящая через линию пересечения плоскостей S_1, S_2 и перпендикулярная плоскости S_3, и еще одна плоскость S_1', такая, что

$$S_2S_1 = S_2'S_1'$$

и

$$F = S_3S_2'S_1'.$$

В полученном представлении автоморфизма F плоскости S_3 и S_2' перпендикулярны. В пучке плоскостей, содержащем плоскости S_2' и S_3, найдется плоскость S_3', перпендикулярная плоскости S_1', и еще одна плоскость S_2'', такая, что

$$S_3S_2' = S_3'S_2''.$$

Заметим, что плоскости S_2'', S_3' взаимно перпендикулярны так же, как и плоскости S_2', S_3. Новое представление автоморфизма F имеет теперь вид:

$$F = S_3' S_2'' S_1',$$

где плоскость S_3' перпендикулярна плоскостям S_1' и S_2''.

Произведение отражений $S_2''S_1'$ совпадает либо со сдвигом, вектор которого параллелен плоскости S_3', либо с поворотом, ось которого (линия пересечения плоскостей S_1', S_2'') перпендикулярна плоскости S_3'. В первом случае отображение называется *переносом с отражением*, во втором — *поворотом с отражением*.

Перенос с отражением мы получим, выполнив сначала сдвиг, а затем отражение в плоскости, которой принадлежит вектор сдвига. При повороте с отражением мы сначала производим поворот, а затем отражение в плоскости, перпендикулярной оси поворота. Отражение можно считать частным случаем и переноса с отражением, и поворота с отражением.

Итак, *движения пространства являются винтовыми движениями. Все остальные автоморфизмы пространства исчерпываются переносами с отражением и поворотами с отражением.*

Является ли данный автоморфизм движением или нет, на плоскости определяется тем, сохраняет ли данный автоморфизм ориентированные углы или изменяет их знак. Существует ли в пространстве аналогичный критерий, позволяющий отличать движения? Прежде чем мы сумеем ответить на этот вопрос, нам понадобится довольно длинное отступление.

Ориентация

Обычно, говоря о пространстве, мы имеем в виду трехмерное пространство, в котором живем. Под *n-мерным пространством* ($n = 1, 2, \ldots$) математик понимает множество элементов, или точек, представимых в виде линейных комбинаций некоторых выделенных элементов, то есть в виде суммы выделенных элементов, умноженных на действительные числа. Число таких выделенных элементов служит важной характеристикой пространства и совпадает с его размерностью. Совокупность всех выделенных элементов называется *базисом* пространства.

Начнем с одномерного пространства — с прямой. Ее структура определяется понятием расстояния между двумя точками. Стоит задать на прямой произвольно выбранный отрезок, приняв его длину за 1, как, умножая его на действительные числа, мы получаем возможность измерять расстояния между любыми двумя точками.

Итак, указав на прямой любые две точки, мы можем сказать, чему равно расстояние между ними. Именно это имеют в виду математики, когда говорят о том, что прямая наделена метрической структурой. Отображениями, сохраняющими метрическую структуру прямой, — *автоморфизмами* этой структуры — служат сдвиги и отражения относительно точек. Они образуют группу — группу автоморфизмов метрической структуры прямой. Сдвиги образуют подгруппу этой группы. Существует ли на прямой нечто такое, что сохранялось бы при сдвигах, но могло бы изменяться при других автоморфизмах.

Да, на прямой существует еще одна, более «первичная» (по сравнению с метрической) структура — так называемая *структура* порядка. Стоя на прямой, можно отличить направление «направо» от направления «налево», или, если представлять себе прямую в виде оси времени, отличить прошлое от будущего. Что называть правым и что левым, безразлично и зависит всецело от нашего вкуса. Но коль скоро выбор произведен, то прямая становится *ориентированной.* Ориентацию можно задавать стрелкой, указывающей направление слева направо. Впрочем, достаточно задать ориентацию базисного (единичного) отрезка, чтобы знать, как ориентированы любые отрезки и вся прямая.

Прямую можно ориентировать двумя способами. При одной ориентации точка a лежит слева от точки b, при другой — точка b лежит слева от точки a. Обе ориентации прямо противоположны.

В чем состоит различие между сдвигами и отражениями на прямой? При сдвигах ориентация прямой сохраняется, при отражениях — изменяется на противоположную.

Если прямая наделена обеими структурами — метрической структурой и структурой близости, то группа автоморфизмов прямой сужается: в ней остаются лишь сдвиги. Группа автоморфизмов метрической прямой содержит сдвиги и отражения относительно точек, группа автоморфизмов метрической ориентированной прямой содержит только сдвиги.

Перейдем к двумерному пространству — плоскости, наделенной обычной метрической структурой (для любой пары точек плоскости расстояние определено, как обычно). Группа автоморфизмов в этом случае состоит из сдвигов, поворотов и переносов с отражениями. Можно ли ввести на плоскости структуру близости?

Стоя на плоскости и раскинув руки в стороны, мы можем наметить прямую, ориентированную, например, слева направо. Двигаясь лишь вдоль этой прямой, мы можем заимствовать для лежаших на ней точек плоскости структуру близости, введенную на прямой. На оси времени нам просто не остается ничего другого: вернуться назад во времени невозможно. На плоскости мы можем сойти с прямой, повернуться на 180° и снова ступить на нее, по-прежнему задавая ориентацию раскинутыми в стороны руками. Очутившись снова на прямой, мы заметим, что ориентация ее изменилась. Но изменилось не только направление «слева направо», но и направление, в котором мы смотрим, глядя прямо перед собой,—направление «вперед — назад». На плоскости понятие «налево — направо» имеет смысл лишь тогда, когда оно связано с понятием «вперед — назад». Связь между тем и другим определяется командами: «Направо!» и «Налево!» По команде «Налево!» мы должны повернуться так, чтобы наша правая рука указывала туда, где раньше было направление «вперед».

В двумерном пространстве — на плоскости — существуют два независимых базисных элемента. Если мы вздумаем переводить один из них в другой, то тем самым какие-то точки плоскости окажутся брошенными на произвол судьбы. Третьего базисного элемента на плоскости не существует.

Ориентировать плоскость означает указать, в какую сторону следует поворачиваться по команде «Направо!» и в какую — по команде «Налево!» Сделать это можно четырьмя различными способами. Прежде всего необходимо задать ориентацию на любой прямой и установить, какая из ее сторон обращена вперед. Тем самым ориентация и направление «вперед» будут заданы для всех прямых. Правда, вернувшись на прямую после прогулки по плоскости, мы можем обнаружить, что направления «вперед» и «назад» поменялись местами так же, как и направления «налево» и «направо», но поворот «налево» останется поворотом «налево».

Ориентацию плоскости можно задать также при помощи двух ориентированных пересекающихся прямых: первая прямая будет указывать направление «слева направо», вторая — направление «вперед» (и тем самым противоположное направление «назад»). Но ориентация плоскости зависит от того, какую из прямых мы сочтем первой и какую — второй. Ориентацию плоскости наглядно можно задать, начертив замкнутую кривую с указанным на ней направлением обхода (рис. 100).

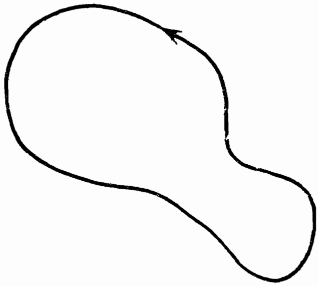


Рис. 100.

Итак, на каждую неориентированную плоскость приходится по две ориентированные. Автоморфизмами неориентированной метризованной плоскости служат сдвиги, повороты и переносы с отражениями. При переходе к ориентированной плоскости переносы с отражениями «вымерзают», поскольку если встать на ось симметрии, то направление «слева направо» при отражении сохраняется, а направление «вперед» изменяется на противоположное. На прямой, перпендикулярной оси симметрии, при отражении изменяется направление «слева направо», но сохраняется направление «вперед». И в том, и в другом случае ориентация не останется неизменной.

Задание ориентации на плоскости — не роскошь. Ориентация нужна хотя бы для того, чтобы знать, в какую сторону крутить ручку кофемолки. Евклид пренебрег ориентацией плоскости. Все свои построения он проводил на неориентированной плоскости, и по традиции его примеру и поныне следуют в школе. И хотя Евклид не имел представления о кофемолках, пренебрежение ориентацией плоскости уже в его времена причиняло известные неудобства при попытках практического применения геометрических теорем. Рассмотрим, например, столь важное понятие, как угол. Евклид всегда считает углы положительными и принимающими значения от 0° до 180° (у Евклида это выражено несколько иначе, поскольку он не знал градусной меры). Для Евклида все прямые углы равны, для астронома и геодезиста они могут быть разными. Если астроном и геодезист отмерят вдоль меридиана дугу, стягивающую прямой центральный угол, но ошибутся при выборе направления, то ничего хорошего из этого не получится. Если за начало отсчета дуги принять экватор, то, пройдя вдоль меридиана дугу в четверть окружности, можно оказаться не на Северном, а на Южном полюсе. Разумеется, во избежание подобных ошибок можно вводить восточную и западную долготу, южную и северную широту, однако это довольно громоздко и не слишком удобно.

Но в один прекрасный день транспортир с дугой в 180° заменили транспортиром с дугой в 360° и сделали вид, будто пользоваться таким транспортиром, содержащим градуированную половину окружности, очень просто. Если требуется измерить угол, то транспортир накладывают на плоскость так, чтобы одна сторона угла (какая именно — безразлично) совместилась с нулевым делением шкалы, вершина угла совпала с центром полуокружности, а другая сторона угла пересекала полуокружность. Отметка шкалы транспортира у точки пересечения соответствует величине измеряемого угла. Располагая транспортиром со шкалой, охватывающей полную окружность, прежде чем приступить к измерениям, мы должны еще выбрать один из двух углов — указать в явном виде, какую из сторон угла считать неподвижной и какую подвижной. «Неподвижная» сторона угла проходит через отметку 0°, «подвижная» указывает отметку шкалы, соответствующую величине угла (вершина угла совмещена с центром окружности). Если стороны угла поменять ролями, то, например, угол 90° превратится в угол 270°, или —90°. Но это еще не все. Если мы случайно перевернем (прозрачный) транспортир на другую сторону, то все углы изменяются точно таким же образом и вместо 90° мы получим 270°. Как избежать такой неприятности? Можно, конечно, воспользоваться цифрами на шкале транспортира: следить за тем, чтобы, например, цифра 2 выглядела, как 2, а не как зеркальное отражение двойки. Но отличить двойку от ее зеркального отражения нам удастся лишь на ориентированной плоскости. Транспортир с круглой шкалой пригоден для использования лишь на ориентированной плоскости.

Впрочем, если говорить о «способе употребления» кругового транспортира, то нельзя не признать, что кое-какие концы у нас все же не сходятся. Пользуясь таким транспортиром, мы отнюдь не находимся в самой плоскости, а накладываем его на плоскость из пространства. С какой стороны мы должны смотреть на транспортир в пространстве?

Так возникает вопрос относительно ориентации трехмерного пространства. Команды «Направо!», «Налево!», помогавшие нам установить ориентацию плоскости, оказываются бесполезными в пространстве так же, как «направление, идущее слева направо», при переходе от прямой к плоскости. Чаще всего мы этого даже не замечаем, поскольку проводим большую часть жизни на плоскости, точнее, на ограниченной части земной поверхности. Команда «Налево!» имеет вполне однозначный смысл, поскольку каждому ясно, что выполнять ее нужно, стоя на ногах, а не на голове. Если кому-нибудь довелось бы выполнить команду «Налево!», стоя на голове, то случайные очевидцы этого необычного зрелища объяснили бы ему, что он повернулся направо. Кто не верит, может испробовать это сам (разумеется, на собственный страх и риск). Чтобы избавить читателя от опасности свернуть шею, предложим ему следующий эксперимент. Нужно взять карандаш, грани которого окрашены в четыре цвета, поставить его перед собой и начать вращать влево. (Если читатель не знает, что такое «влево», мы советуем ему зажать карандаш между ладонями и вращать, двигая правую руку вперед, а левую назад. Где правая рука, а где — левая, знает каждый.)

Перед нами одна за другой будут проходить разноцветные грани: красная, черная, зеленая, синяя. Если поставить карандаш на стол другим концом и снова начать вращать его, то последовательность, с которой станут сменяться грани, будет противоположной. Оба раза мы крутили карандаш в левую сторону, но сам карандаш, владей он даром речи, сказал бы, что один раз его крутили влево, а другой — вправо. Вообразим теперь, что перед нами не карандаш, а танцор. Если он, вальсируя, делает повороты в левую сторону, то мы увидим (по порядку) его лицо, правое ухо, спину и левое ухо. Если бы танцор проделал те же па, стоя на голове, то мы увидели бы сначала его левое ухо, потом спину, правое ухо и лицо.

К двум выделенным базисным элементам на плоскости в трехмерном пространстве добавляется еще один, третий, элемент. Наше пространство можно «натянуть» на три и только три выделенных направления.

Что означают в пространстве выражения «направо» и «налево», зависит еще и от того, как понимать выражения «вверх» и «вниз». В большинстве случаев ясно, где верх и где низ, и поворачиваться налево нам приходится, стоя не на голове. Ручку кофейной мельницы согласно приложенной инструкции надлежит крутить вправо, а о том, чтобы перевернуть кофейную мельницу «вверх тормашками» не может быть и речи: из нее высыпались бы все кофейные зерна. Гораздо труднее договориться о том, где правая сторона и где левая, двум людям, если им нужно привести во вращение слева направо колесо или точильный камень. «Правое» и «левое» в пространстве имеют смысл лишь тогда, когда они связаны с понятиями «верх» и «низ». Людям необходимо условиться о том, какую сторону плоскости колеса называть верхней и какую — нижней. Необходимо также, как говорят еще, задать ориентацию оси — *вектор*, направленный снизу вверх. Если *вектор оси* не задан, то понятия «правое» и «левое» утрачивают смысл. На оконном стекле написали цифру 2. Снаружи или изнутри? Ее верхняя часть — «лебединая шея» — изогнута вправо, если смотреть... (Как тут не вспомнить точные указания относительно того, как пользоваться круглым транспортиром.)

Связь «правого» и «левого» с понятиями «верх» и «низ», направления вращения колеса с вектором оси приводят к ориентации пространства. Существуют реальные предметы, в которых эта связь воплощена в металле: винты, штопоры и буравы. *Перед* нами находится *правый винт*, если нити резьбы идут *слева и снизу направо и вверх*. Это определение не зависит от нашего расположения относительно винта. Мы можем обойти винт и встать с противоположной стороны его, можем встать на голову. (Лучше все-таки перевернуть винт!) Правый винт — это комбинация движения слева направо и внутрь (вниз) или справа налево и наружу (вверх). Стоит нам заменить движение слева направо движением справа налево *или* движение вниз движением вверх, как правый винт перейдет в левый. Если же мы одновременно произведем обе замены, то правый винт останется правым винтом. Говоря о вращающейся вокруг своей оси Земле или о вращающемся электроне, нельзя сказать, в какую сторону они крутятся: направо или налево. Ответ зависит от того, куда мы направим ось вращения. Земля, вращающаяся вокруг своей оси, не ориентирована в пространстве. Но стоит нам присоединить к суточному вращению Земли вектор оси (выбрав его направление от Южного полюса к Северному) и добавить, что «винт должен быть левым», как наше пространство станет ориентированным.

Задавая ориентацию на прямой, мы выбираем на ней две точки и говорим, указывая на одну из них: «Эта лежит слева». Вводя ориентацию на плоскости, мы отдаем предпочтение определенному направлению вращения, когда объясняем, как следует выполнять поворот «налево». Пространство обретает ориентацию, когда мы указываем винт, который называем левым винтом. Вместо винта можно было бы взять ориентированную плоскость, указав, какая из ее сторон считается верхней и какая нижней (положительной или отрицательной), или проткнув ее каким-нибудь вектором. Ориентацию пространства можно увидеть наглядно, задав три взаимно перпендикулярные ориентированные прямые x_1, x_2, x_3. В плоскости x_1x_2 ориентация задается направлением обхода точки пересечения прямых x_1 и x_2, при котором от оси x_1

к оси x_2 можно перейти, описав угол 90°. Направление оси x_3 выбирают за направление вектора оси. Для установления ориентации пространства можно воспользоваться волчком. Говоря о волчке, мы имеем в виду круглый диск с ориентированным краем и с перпендикулярной его плоскости стрелкой в центре (рис. 101). Узлы на веревке также задают ориентацию пространства: чтобы убедиться в этом, достаточно обратить внимание на винтовое движение, которым мы начинаем завязывать узел («правши» предпочитают описывать концом веревки правый винт). Но всякий раз, когда мы строим в пространстве «винт», необходимо добавлять: «Этот винт мы называем правым или положительным». (Некоторые называют положительным левый винт.)

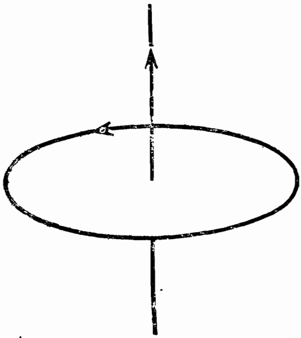


Рис. 101.

На каждое неориентированное метрическое пространство приходятся два ориентированных. То, что в одном из них задает левый винт, в другом определяет правый.

Автоморфизмами обычного метрического пространства служат винтовые движения, переносы с отражением и повороты с отражением. В ориентированном пространстве из автоморфизмов остаются лишь винтовые движения. При отражении в зеркале правый винт переходит в левый (если ось винта лежит в плоскости зеркала, то из-за изменения на противоположное направление поворота; если ось винта перпендикулярна плоскости зеркала, то из-за опрокидывания вектора оси).

Какое пространство, неориентированное или ориентированное, необходимо для *описания* явлений *природы*?

Можно не сомневаться: нам совершенно необходимо ориентированное пространство. В неориентированном пространстве мы не можем воспользоваться даже штопором, чтобы вытащить пробку из бутылки.

Физик также нуждается в ориентированном пространстве, ибо в противном случае как бы он стал закручивать гайки или винты, монтируя свои приборы? Биохимик также не мог бы обойтись без ориентированного пространства, если бы ему потребовалось отличать природные левовращательные соединения от симметричных им правовращательных соединений.

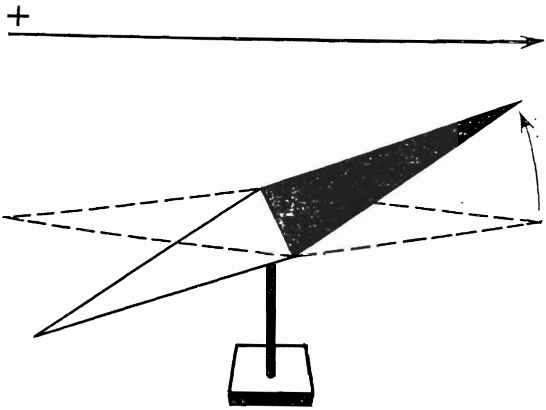


Рис. 102.

Не является ли все это нагромождением случайностей? Разве нельзя было бы навить штопор или нарезать винт в обратную сторону? Разве не могла бы развиваться жизнь, если бы природные оптические активные соединения были правовращательными?

В природе де-факто существует один тип винта. Но существует ли он де-юре? Необходимо ли ориентированное пространство для формулировки общих законов природы? В механике понятие ориентированного пространства заведомо никак не используется, а в электромагнетизме? В теории электромагнитных явлений встречается так называемое *правило буравчика*. Если по проводнику, натянутому над (или под) магнитной стрелкой, пропускать ток (рис. 102), то магнитная стрелка отклоняется влево или вправо. Правило буравчика позволяет предсказывать заранее, в какую сторону отклонится стрелка. Оно гласит: если направить ось (правого) штопора вдоль проводника и мысленно ввинчивать его в направлении (условно принимаемом за направление) тока, то северный полюс магнитной стрелки отклонится в ту сторону, в какую его толкнула бы ручка штопора. Означает ли это признание правого винта де-юре в качестве законов природы? На первый взгляд кажется, что означает. Если мы забыли, какой винт заслуживает названия правого, то нужно спросить об этом у природы и дождаться ответа. Проведем такой эксперимент. Пропустим ток по проводнику, натянутому над магнитной стрелкой, и по ее отклонению установим, какой из двух типов штопора служит моделью правого винта. Казалось бы, сама природа выделяет определенный тип винтового движения. Но не будем торопиться с выводами.

Прежде чем прийти к заключению о «предвзятости» природы к винту, называемому нами правым, необходимо ответить на два вопроса. Во-первых, в каком направлении течет электрический ток? Во-вторых, какой из двух концов магнитной стрелки северный? И направление тока, и северный конец магнитной стрелки выбраны произвольно, но вполне определенным образом. Электрический ток в батарейке от карманного фонаря течет от цинкового электрода к угольному. Следовательно, чтобы установить направление, в котором течет электрический ток, нужно лишь вскрыть батарейку. Северный конец магнитной стрелки показывает на север, по крайней мере в большинстве мест на земном шаре. Если мы хотим определить, какой из двух концов магнитной стрелки северный, то нам необходимо знать направление на север. Вдали от Земли это правило нам не поможет. Если мы отправимся в космическое путешествие к дальним мирам и по дороге вздумаем задавать природе вопрос о том, какой из двух полюсов магнитной стрелки должен считаться северным, то, хотя ответ нам и будет дан, истолковать его все равно будет невозможно. У нас, на Земле, принято окрашивать северный конец магнитной стрелки в синий цвет, а южный — в красный. Захватив с собой такую сине-красную стрелку, мы сможем всюду пользоваться правилом винта, чтобы при помощи его отличать правый винт от левого.

Но вот мы захотели рассказать физикам, обитающим на дальних мирах о том, что такое правый винт. Нам удалось установить с ними двустороннюю радиосвязь. Мы научились переводить свою, земную, физику на язык их небесной физики, и наоборот, и теперь жаждем объяснить нашим коллегам, как движется правый винт. Объяснить, дав строгое теоретическое определение, а не просто продемонстрировав правый винт. Мы не можем поэтому указать на какие-нибудь звезды, случайно образовавшие правый винт. Мы не можем послать инопланетным физикам циркулярно-поляризованный свет с объяснением, что это — право-поляризованный свет. Что же делать? Правило буравчика также нельзя использовать, поскольку мы не в силах объяснить, какой конец магнитной стрелки северный.

Повторим еще раз: знать какой винт правый и какой левый, нам необходимо для того, чтобы мы могли открыть бутылку вина, отвернуть гайку, знать, в какую сторону вращается небосвод, проводить эксперименты с током и магнитной иглой, но направление винта в конечном счете можно лишь показать, а не определить теоретически. Лучше всего это можно понять, если попытаться объяснить (а не показать) физикам, живущим на дальних окраинах Вселенной, что такое правый винт.

Так или примерно так думали еще несколько лет назад. Непризнание правого винта де-юре было догмой, в которой не сомневался ни один физик. Но в 1956 г. произошло событие, которое по праву можно назвать самым значительным вкладом в философию пространства со времен создания теории относительности: удалось обнаружить в природе магниты с помеченными северными полюсами. Разумеется, эти магниты не были столь неуклюжими, как стрелки наших компасов, и северные полюсы их не были закрашены в синий цвет.

Магниты, которые мы имеем в виду,—это радиоактивные ядра атомов. Атомное ядро обычно ведет себя, как волчок. Ядро несет положительный заряд, который, вращаясь, создает круговой ток и превращает ядро в магнит. Но волчок сам по себе, без вектора оси, не определяет направление винта. А магнит без помеченного природой северного полюса для наших целей бесполезен.

Если взять не стабильные, а β-радиоактивные ядра, то они могут испускать электроны. Направления, в которых электроны вылетают из ядер, можно использовать вместо недостающего вектора оси, по крайней мере, если они не все перпендикулярны оси магнита и не равнораспределены между верхом и низом.

Эксперименты, проводившиеся на протяжении восьми лет с образцом ^{60}Co, показали, что существует предпочтительное направление испускания электронов, образующее с направлением вращения ядра левый винт. Иначе говоря, магниты — ядра ^{60}Co несут на своих южных полюсах особую отметку, которая гласит: «Преимущественное направление испускания электронов проходит здесь».

На помеченные природой магниты экспериментаторы наткнулись не случайно, и помогла найти их не слепая удача. Американские физики Янг и Ли после тщательного анализа мезонных распадов высказали сомнение в правильности общепринятой догмы, согласно которой ориентацию пространства нельзя определить чисто теоретическим путем. Свои сомнения они облекли в форму эксперимента, который и был осуществлен по их предложению. Сейчас, много лет спустя после проведения эксперимента, полученный результат можно объяснить на общедоступном языке. Но те глубокие теоретические соображения, которые привели к нему, исторически были вполне оправданы, поскольку предсказанный Ли и Янгом исход эксперимента опрокидывал все привычные представления.

Если мы захотим сообщить де-юре физикам, находящимся в любой точке Вселенной, что такое правый винт, то нам достаточно отправить телеграмму следующего содержания: «В β-радиоактивных ядрах ^{60}Co направление вращения ядер образует с направлением преимущественного испускания электронов левый винт». Повторив эксперимент, предложенный Ли и Янгом, инопланетные физики смогут, комбинируя результаты измерений с правилом буравчика, понять смысл таких земных выражений, как «направо» и «налево». Мы получаем возможность рассказать (а не показать) обитателям далеких миров, что такое правое и левое. Правда, наша затея была бы обречена на провал, если бы...

Наша затея заведомо была бы обречена на провал, если бы ядра ^{60}Co у инопланетян были не левыми, а правыми винтами. Такую возможность нельзя полностью исключить. Мир, в котором живут наши ученые коллеги, может оказаться антимиром. Известно, что наш реальный мир устроен весьма несимметрично. Все наши электроны (легкие частицы) несут отрицательные заряды. Положительно заряженные частицы встречаются в нашем мире очень редко, они не могут долго храниться у нас, поскольку при столкновениях с отрицательно заряженными и те и другие аннигилируют. Наши более тяжелые заряженные частицы (протоны) почти все несут положительный заряд. Но и они, встречая своих более редких отрицательных коллег, аннигилируют вместе с ними. Уже давно высказывалось подозрение, что избыток отрицательно заряженных электронов и положительно заряженных протонов служит своего рода локальной характеристикой того участка Вселенной, где мы обитаем. Где-нибудь в другом месте роль привычных нам частиц могли бы исполнять античастицы, которые выше мы назвали просто «частицами». Граница между миром и антимиром стала бы грандиозным полем битвы между протонами и электронами. В антимире протон в ядре был бы заряжен отрицательно, а электрон — положительно. Показания ядерных магнитов ^{60}Co оказались бы «неверными». В среднем ядре анти-^{60}Co определяли бы правый, а не левый винт. Можно было бы думать, что специалист по антиэлектромагнетизму интерпретировал бы наше правило буравчика в смысле, обратном принятому в нашем мире, поскольку в его антибатарейке от карманного фонарика электроды сделаны из антицинка и антиуглерода и ток идет в обратном направлении. Не следует, однако, упускать из виду, что и магниты его сделаны из антижелеза и поэтому обращают весь эффект еще раз. Если бы наши физики захотели рассказать антифизикам, что такое правый винт, то те сочли бы их ограниченными провинциалами. Язык, на котором изъяснялись бы наши физики, в ушах их антиколлег звучал бы как некий непонятный обитателям антимира диалект, непонятный по крайней мере до тех пор, пока антимиряне не узнали бы, что послание пришло к ним из «антимира» (по отношению к их миру).

Вопрос, которому мы уделили достаточно внимания, можно теперь сформулировать более точно: можно ли в нашем пространстве предложить правило, позволяющее чисто словесно определить, что такое правый и что такое левый винт? Наше определение, разумеется, не должно противоречить в каждом отдельном случае результатам повседневного, нетеоретического, основанного на показаниях приборов опыта. Ответ на этот вопрос после обнаружения возможности задать правый винт де-юре, как мы видели, зависит от структуры де-факто Вселенной (которая, быть может, в конце концов окажется структурой де-юре).

Оглавление

Предисловие переводчика 5

Введение 8

Глава 1. Человек измеряет Вселенную 15

Глава 2. До бесконечности 49

Глава 3. Чем занимаются вычислительные машины 73

Глава 4. Азбука жизни 133

Глава 5. Искусство рисовать плохо 149

Глава 6. „Дайте мне точку опоры и я переверну Землю" 175

Глава 7. Мир в зеркале 207