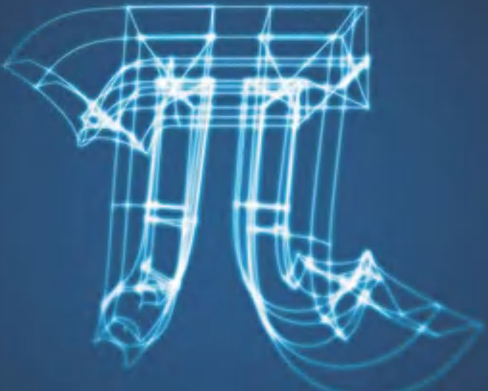
ИЭН СТЮАРТ



ВЕЛИЧАЙШИЕ МАТЕМАТИЧЕСКИЕ ЗАДАЧИ

Выдающийся и наиболее продуктивный среди британских популяризаторов математической науки The Guardian

Династия

анф

ВЕЛИЧАЙШИЕ МАТЕМАТИЧЕСКИЕ ЗАДАЧИ

THE GREAT MATHEMATICAL PROBLEMS

Marvels and Mysteries of Mathematics

IAN STEWART

PROFILE BOOKS

ВЕЛИЧАЙШИЕ МАТЕМАТИЧЕСКИЕ ЗАДАЧИ

ИЭН СТЮАРТ

Перевод с английского

Династия
анф

Москва
2016

УДК 519.6
ББК 22.1
С88

Переводчик Наталья Лисова
Редактор Наталья Нарциссова

Стюарт И.
С88 Величайшие математические задачи / Иэн Стюарт ; Пер. с англ. — М.: Альпина нон-фикшн, 2016. — 460 с.

ISBN 978-5-91671-318-3

Закономерности простых чисел и теорема Ферма, гипотеза Пуанкаре и сферическая симметрия Кеплера, загадка числа π и орбитальный хаос в небесной механике. Многие из нас лишь краем уха слышали о таинственных и непостижимых загадках современной математики. Между тем, как ни парадоксально, фундаментальная цель этой науки — раскрывать внутреннюю простоту самых сложных вопросов. Английский математик и популяризатор науки, профессор Иэн Стюарт, помогает читателю преодолеть психологический барьер. Увлекательно и доступно он рассказывает о самых трудных задачах, над которыми бились и продолжают биться величайшие умы, об истоках таких проблем, о том, почему они так важны и какое место занимают в общем контексте математики и естественных наук. Эта книга – проводник в удивительный и загадочный мир чисел, теорем и гипотез, на передний край математической науки, которая новыми методами пытается разрешить задачи, поставленные перед ней тысячелетия назад.

УДК 519.6
ББК 22.1

ISBN 978-5-91671-318-3 (рус.)
ISBN 978-1-84668-1998 (англ.)

Династия

Фонд некоммерческих программ
«Династия»
основан в 2002 г.
Дмитрием Борисовичем Зиминым,
почетным президентом компании «Вымпелком».
Приоритетные направления деятельности Фонда —
поддержка фундаментальной науки и образования в России,
популяризация науки и просвещение.
В рамках программы по популяризации науки
Фондом запущено несколько проектов.
В их числе — сайт elementy.ru, ставший одним
из ведущих в русскоязычном Интернете тематических ресурсов,
а также проект «Библиотека «Династии» —
издание современных научно-популярных книг,
тщательно отобранных экспертами-учеными.
Книга, которую вы держите в руках, выпущена
в рамках этого проекта.
Более подробную информацию о Фонде «Династия»
вы найдете по адресу
www.dynastyfdn.ru.

Содержание

Мы должны знать — мы будем знать!

Давид Гильберт,
речь о математических проблемах, произнесенная в 1930 г.
по случаю присвоения Гильберту
звания Почетного гражданина Кёнигсберга

Предисловие

Математика — обширная, непрерывно растущая и столь же непрерывно меняющаяся область знания. Среди бесчисленных вопросов, которыми задаются математики и на которые они по большей части находят ответы, есть немало и таких, которые стоят особняком и возвышаются над всеми прочими, словно горные пики — над предгорьями. Это действительно сложные проблемы, и любой математик отдал бы правую руку за возможность первым найти решение одной из таких масштабных задач. Некоторые из них оставались нерешенными десятилетиями, иные — столетиями, а есть и такие, что не поддавались усилиям математиков несколько тысячелетий. И до сих пор существуют проблемы, которые ученым только предстоит разрешить. Так, последняя теорема Ферма оставалась для математиков камнем преткновения 350 лет, пока Эндрю Уайлс не доказал ее, потратив на эту работу семь лет жизни. Гипотеза Пуанкаре была неприступна больше 100 лет, пока эксцентричный гений Григорий Перельман не нашел доказательство и не превратил ее в теорему (отказавшись при этом от всяких академических почестей и премии в миллион долларов за эту работу). А гипотеза Римана и сегодня, через 150 лет после того, как была сформулирована, остается нерешенной.

Книга «Великие математические задачи» рассказывает о некоторых крупнейших математических проблемах, работа над которыми открыла перед научной мыслью совершенно новые направления и возможности. Читатель познакомится с истоками этих задач, узнает, почему они так важны и какое место занимают в общем контексте математики и естественных наук. В книге представлены как решенные, так и нерешенные задачи из самых разных периодов истории математики.

По существу, рассказ охватывает две с лишним тысячи лет развития науки, однако основное внимание в книге сосредоточено на вопросах, которые либо до сих пор остаются нерешенными, либо были решены относительно недавно, в последние полвека.

Фундаментальная цель математики — раскрывать внутреннюю простоту сложных на первый взгляд вопросов. Это заявление может показаться неочевидным и даже странным, поскольку математическое представление о «простоте» опирается на множество сложных технических концепций. Но важная особенность этой книги заключается именно в том, что акцент в ней сделан на глубинную простоту, а сложности мы стараемся обойти стороной или объясняем простыми словами.

Математика — более молодая и многообразная наука, чем многие думают. По приблизительным оценкам в мире сегодня живет около 100 000 математиков-исследователей, которые каждый год выпускают более *двух миллионов* страниц новых математических изысканий. Это не «новые числа», поисками которых математики не занимаются вообще. И не «новые величины», подобные уже известным, только больше, хотя мы действительно иногда работаем с достаточно большими величинами. Так, про одно недавнее алгебраическое исследование, проведенное командой из 25 математиков, какой-то шутник сказал: «Расчет размером с Манхэттен». Но и это не совсем верно — ребята поскромничали. Размером с Манхэттен у них был *ответ,* а расчет занимал гораздо больше места. Впечатляет, не правда ли? Но главное в математических исследованиях все-таки качество, а не размер и даже не количество. Расчет размером с Манхэттен, о котором шла речь, котируется в обоих отношениях, поскольку дает важную информацию о группе симметрии, играющей существенную роль в математике и, судя по всему, в квантовой физике. Блестящие математические рассуждения и выводы могут уложиться в одну строчку — а могут занять целую энциклопедию. Все зависит от существа и сложности задачи.

При мысли о математике на ум в первую очередь приходят страницы, заполненные малопонятными значками и формулами. Однако те два миллиона страниц, о которых мы говорили, содержат по большей части слова, а не специальные символы. Слова необходимы для объяснения существа проблемы, описания хода мысли и смысла вычислений, и, кроме того, без них невозможно объяснить, какое место все это занимает в постоянно строящемся здании математики. Как заметил на границе XVIII и XIX вв. великий Карл Гаусс, главное в математике — «идеи, а не символы». Тем не менее обычно математические идеи излагаются языком символов. И многие исследовательские работы содержат больше символов, чем слов. Четкости, которую обеспечивают формулы, не всегда можно достичь словами.

Тем не менее нередко математические идеи можно объяснить словами, оставив в стороне большую часть специальных символов. И именно этот принцип лег в основу книги, которую вы держите в руках. Она рассказывает, чем занимаются математики, как они думают и почему предмет их исследований интересен и важен для всего человечества. Она показывает также (и это очень важно), как сегодняшние математики справляются с вызовами своих предшественников, как одна за другой великие загадки прошлого уступают мощным методикам настоящего, тем самым изменяя и математику, и естественные науки будущего. Математика по праву относится к величайшим достижениям человечества, и ее важнейшие задачи — решенные и нерешенные — уже не одну тысячу лет направляют и стимулируют творческие силы человека.

Ковентри, июнь 2012 г.

1

Великие задачи

Телепередачи о математике попадаются редко, а хорошие и того реже. Одной из наиболее удачных среди них, причем не только по содержанию, но и по степени увлекательности и вовлеченности зрителей, стала программа о Великой теореме Ферма, которую в 1996 г. снял для научно-популярной серии Horizon британской корпорации BBC Джон Линч. Саймон Сингх, который также участвовал в создании этой программы, превратил рассказанную в ней историю в захватывающую книгу-бестселлер. На своем сайте он рассказал, что поразительный успех передачи стал для всех сюрпризом.

> «В нашей программе целых 50 минут математики рассказывают о математике. Не сказать, чтобы это был надежный рецепт создания телевизионного блокбастера, но наша передача взбудоражила зрителей и привлекла внимание критиков. Она получила премию BAFTA как лучшая документальная программа, Приз Италии, другие международные награды и была номинирована на Emmy. Это доказывает, что математика может быть не менее захватывающей темой, чем любая другая».

Я думаю, что успех телепрограммы и книги был обусловлен несколькими причинами, которые имеют немаловажное значение и для моего рассказа. Но чтобы не слишком разбрасываться, я буду говорить только о документальном фильме.

Последняя теорема Ферма — одна из величайших математических проблем, но возникла она из невинного на первый взгляд замечания, сделанного одним из ведущих математиков XVII в. на полях классического учебника. Постепенно проблема приобрела известность, поскольку никто не мог ни доказать, ни опровергнуть утверждение, содержавшееся в оставленной Пьером Ферма заметке на полях. Несмотря на усилия, предпринимавшиеся множеством необычайно умных людей, такое положение вещей сохранялось более 300 лет, поэтому когда в 1995 г. британскому математику Эндрю Уайлсу удалось наконец справиться с этой проблемой, масштаб его достижения был очевиден каждому. Не нужно было даже знать, в чем заключается проблема, не говоря уже о ее решении. В какой-то мере достижение Уайлса — то же самое, что покорение Эвереста.

Помимо научного значения, успешное доказательство теоремы Ферма связано с интереснейшей жизненной историей. В 10 лет Эндрю Уайлс так заинтересовался этой проблемой, что решил стать математиком и обязательно решить ее. Он выполнил первую часть плана и даже выбрал своей специализацией теорию чисел — обширную область математики, к которой относится и Великая теорема Ферма. Однако чем больше он узнавал о математике, тем труднее казалось выполнить задуманное. Теорема Ферма — загадочная диковинка, обособленный вопрос из разряда тех, которые умеет задавать любой специалист по теории чисел (ведь для этого не нужно никаких доказательств). Она не укладывается ни в одну систему мощных доказательных средств. Великий Гаусс в письме к Генриху Ольберсу попросту отмахнулся от нее, заметив, что эта проблема «мне не особенно интересна, поскольку легко можно сформулировать множество подобных утверждений, которые никто не может ни доказать, ни опровергнуть». Уайлс решил, что его детская мечта неосуществима, и отложил теорему Ферма в долгий ящик. Однако затем, будто по волшебству, другие математики совершили прорывное открытие, неожиданно связавшее теорему со стержневой темой теории чисел, причем именно той, которой и занимался Уайлс. Гаусс, как оказалось, в свое время недооценил значение этой проблемы, что для него вообще-то было нехарактерно; он не подозревал, что она может быть связана с глубокой, но на первый взгляд достаточно далекой областью математики.

Теперь, когда связь была установлена, Уайлс мог работать над загадкой Ферма и одновременно проводить значимые исследования в рамках современной теории чисел. Даже если с доказательством Великой теоремы ничего бы не получилось, все, что удалось открыть в ходе исследований, было бы достойно публикации. Так что старые наработки были извлечены на свет божий, и Уайлс начал всерьез обдумывать проблему. Через семь лет усердных трудов (а работал он втайне от ученого сообщества, что для математиков совсем не характерно) Уайлс пришел к выводу, что решение найдено. На престижной конференции по теории чисел он прочел серию лекций под невнятным названием, которое никого не обмануло. Новость разлетелась и произвела сенсацию, причем не только в академических кругах, но и в средствах массовой информации. Теорема Ферма доказана!

Полученное Уайлсом доказательство, полное оригинальных идей, оказалось красивым и элегантным. К несчастью, специалисты вскоре обнаружили в его логике серьезный пробел. Как это ни печально, при решении великих (и обычно очень известных) математических задач такое происходит сплошь и рядом, и, как правило, для очередного доказательства такой поворот событий оказывается роковым. Однако на этот раз судьба была благосклонна: при помощи бывшего своего ученика Ричарда Тейлора Уайлсу удалось ликвидировать пробел, исправить доказательство и завершить работу. Эмоциональное напряжение этого момента очень хорошо видно на экране: пожалуй, это единственный случай, когда ученый-математик расплакался перед камерой при одном только воспоминании о тех драматических событиях и последовавшем за ними триумфе.

Вы, наверное, заметили, что я так и не рассказал вам, в чем, собственно, *заключается* Великая теорема Ферма. Я сделал (или, вернее, не сделал) это намеренно: о самой теореме речь пойдет в свое время. Ведь успех телепередачи с сутью теоремы почти не связан. Мало того, математики никогда не придавали особого значения тому, верна ли теорема, которую Ферма небрежно набросал на полях книги, или нет. От ответа на этот вопрос ничего особенно важного не зависит. Откуда же такой интерес к нему? Все очень просто. Огромное значение может иметь именно то, что все математическое сообщество было не в состоянии *найти* этот ответ. И дело вовсе не в самоуважении: это означало, что в существующих математических теориях не хватает чего-то принципиально важного. К тому же теорема очень просто формулируется, и это добавляет загадочности всей ситуации. Как может что-то настолько на первый взгляд простое оказаться таким сложным?

Математиков не слишком заботил ответ на вопрос, поставленный Ферма, зато глубоко заботил тот факт, что они ответа не знают. К тому же им хотелось найти метод решения этой проблемы, поскольку он, по идее, должен был пролить свет не только на вопрос Ферма, но и на множество других вопросов. Опять же так нередко случается с математическими загадками: методы, использованные для их решения, часто важнее результатов. Разумеется, иногда результат тоже важен — все зависит от его следствий.

Доказательство Уайлса слишком сложно для телепередачи, разобраться в нем могут только специалисты. В нем есть математическая красота и интрига, как мы убедимся в свое время, но любая попытка объяснить что-то подобное по телевизору привела бы к немедленной потере интереса у большей части аудитории. Поэтому программа разумно сосредоточилась на более личном вопросе: каково это — решить математическую проблему, известную своей сложностью и влекущую за собой целый шлейф исторических ассоциаций? Телезрителям показали, что существует небольшая, но увлеченная группа математиков, разбросанных по всему миру, что все они глубоко погружены в предмет своих исследований, общаются друг с другом, следят за последними разработками и вообще посвящают значительную часть жизни продвижению математических знаний. Создатели фильма очень живо показали эмоциональную вовлеченность и социальное единство этих людей. Это не разумные автоматы, а реальные люди, любящие свое дело. В этом и заключается главный посыл фильма.

Мы можем сформулировать три основные причины успеха этой программы: серьезная и известная проблема, герой с увлекательной, по-человечески интересной историей и группа поддержки — целая каста эмоционально вовлеченных в процесс людей. Но я подозреваю, что существует и четвертая причина, не столь явная. Люди, не связанные с математикой, по многим объективным причинам редко слышат о новых достижениях в этой области, да и не так уж сильно интересуются этим. В газетах лишь изредка упоминается что-нибудь связанное с математикой, а если и упоминается, то лишь приводятся какие-то отрывочные или тривиальные факты. Наконец, действия и достижения математиков где-то там за кулисами не оказывают, на первый взгляд, никакого влияния на повседневную жизнь. А школьная математика зачастую предстает перед учащимися как уже закрытая книга, где на каждый вопрос есть готовый ответ. Школьникам обычно кажется, что ничего нового в математике днем с огнем не сыщешь.

Если смотреть под таким углом зрения, то главное в достижении Уайлса — не то, что Великая теорема Ферма была доказана, а то, что наконец-то *в математике свершилось хоть что-то новое*. Поскольку на поиск доказательства теоремы у ученых ушло больше 300 лет, многие зрители восприняли открытие Уайлса как первое существенное достижение в математике за весь этот период. Я не говорю, что все *действительно* именно так и решили. Понятно, что подобная позиция рассыпалась бы в прах при первом же очевидном вопросе вроде: «Почему правительство тратит немалые деньги на финансирование университетских математических исследований?» Но на подсознательном уровне все сочли, что это именно так, не задаваясь вопросами и не размышляя. Поэтому достижение Уайлса приобрело в глазах нематематиков еще большие масштабы.

Одна из целей этой книги — наглядно продемонстрировать всем, в том числе и неспециалистам, что математика сейчас на подъеме, а новые открытия в ней — совсем не редкость. Вы почти ничего об этом не слышите просто потому, что большая часть математических работ слишком сложна для неспециалистов, а средства массовой информации с опаской относятся к интеллектуалам и боятся публиковать что-либо сложнее «Х-фактора». Кроме того, практическое приложение математики обычно скрыто от глаз потребителя, причем зачастую намеренно, чтобы не волновать его. «Что? Работа моего айфона построена на математических формулах? Но у меня же по математике всегда была пара! Как я буду входить в “Фейсбук”?»

Исторически новые достижения в математике часто следуют за открытиями в других областях знания. Исаак Ньютон, разработав законы механики и всемирного тяготения, которые описывают движение планет, не избавился разом от всех проблем в понимании устройства Солнечной системы. Наоборот, после этого перед математиками встал ряд новых вопросов: да, конечно, мы знаем законы, но что они подразумевают? В поисках ответов Ньютон придумал дифференциальное (интегральное) исчисление, но и у нового метода обнаружились ограничения. Зачастую он вместо ответа на вопрос просто дает иную его формулировку. Так, с его помощью некоторые задачи можно легко записать в виде специальной формулы, известной как дифференциальное уравнение. *Решение* этого уравнения и есть искомый ответ. Но это решение еще надо найти. Тем не менее дифференциальное исчисление послужило мощным стартом. Оно показало, что ответ в принципе возможен, и снабдило ученых эффективным методом его поиска. До сих пор, хотя прошло уже больше 300 лет, этот метод помогает математикам совершать крупные открытия.

По мере того как росла сумма математических знаний человечества, все большую роль в мотивации новых исследований стал играть еще один фактор: внутренние запросы самой математики. Если, к примеру, вы умеете решать алгебраические уравнения первой, второй, третьей и четвертой степеней, вам не нужно обладать очень уж богатым воображением, чтобы задаться вопросом об уравнениях пятой степени. (По существу, степень уравнения есть мера его сложности, но чтобы задать очевидный вопрос, не обязательно даже знать, что это такое.) Если решение не дается — как, собственно, и было, — то этот факт *сам по себе* заставляет математиков еще более усердно искать его, и при этом неважно, будет ли вожделенный результат иметь какую-либо практическую пользу.

Я не утверждаю, что практическое приложение не имеет значения. Но если какая-то конкретная математическая составляющая раз за разом возникает в вопросах, скажем, физики волн — океанских волн, вибраций, звука, света, — то понятно, что исследовать ее закономерности было бы полезно. Не обязательно знать заранее, какое приложение найдет новая идея: тема волн фигурирует во многих важных областях, так что серьезные результаты непременно где-нибудь пригодятся. В данном случае этим «где-нибудь» стали радио, телевидение и радары. Если кто-то придумает новый подход к тепловым потокам и без всякого математического обоснования предложит новый блестящий метод, то, безусловно, будет очень полезно разобраться во всем этом *как в чисто математической задаче*. И даже если вам нет никакого дела до тепловых потоков, результат обязательно пригодится где-то еще. Фурье-анализ, разработанный в ходе исследования именно этой области, оказался, возможно, самой полезной математической идеей всех времен. Это, по существу, основа современных телекоммуникаций: он обеспечивает работу цифровых камер, помогает реставрировать старые кинофильмы и звукозаписи, а его современное расширение использует ФБР для хранения отпечатков пальцев.

За несколько тысячелетий подобная взаимосвязь между практическим применением математики и ее внутренней структурой привела к тому, что они тесно переплелись и стали почти неотделимы друг от друга. Тем не менее математика делится на две области: чистую и прикладную. Это деление помогает оценить место математических открытий в структуре человеческого знания, однако оно довольно условно. В лучшем случае так можно различить два конца одного непрерывного спектра математических стилей и методов. В худшем — такая классификация вводит нас в заблуждение относительно того, что именно приносит пользу и что служит источником идей. Как и в других областях науки, силу математике придает сочетание абстрактных рассуждений и вдохновения, почерпнутого из внешнего мира. Говоря попросту, они питают друг друга. Разделить математику на две составляющие не просто невозможно — это бессмысленно.

Большинство по-настоящему важных математических задач — великих задач, которым посвящена эта книга, — возникли внутри математического поля в процессе своеобразной интеллектуальной медитации. Причина проста: это сугубо *математические* задачи. Математика часто представляется набором изолированных областей, в каждой из которых господствуют собственные методы: это алгебра, геометрия, тригонометрия, математический анализ, комбинаторика, теория вероятностей. Ее обычно так и преподают, и не без причины: четкое разделение тем помогает учащимся разложить по полочкам учебный материал в своей голове. И действительно, такое деление — вполне разумный способ понять в первом приближении структуру математической науки, особенно классической, давно устоявшейся. Однако на переднем крае исследований это четкое деление часто рушится. И дело не только в том, что границы между основными областями математики размыты, — в реальности их просто нет.

Каждый математик-исследователь знает, что в любой момент внезапно и непредсказуемо может оказаться, что проблема, над которой он работает, требует свежих идей из какой-то совершенно посторонней, на первый взгляд, области. Более того, новые исследования часто захватывают сразу несколько областей. К примеру, мои исследования сосредоточены по большей части на формировании структур в динамических системах — системах, которые изменяются во времени по определенным правилам. Типичный пример — движение животных. Лошадь при движении рысью раз за разом повторяет одну и ту же последовательность движений ног, и в этих движениях есть четкая закономерность: копыта ударяют по земле попеременно, диагональными парами. Иными словами, лошадь ставит сначала левую переднюю и правую заднюю ноги, затем правую переднюю и левую заднюю. О чем же эта задача? О паттернах, и тогда решать ее надо методами теории групп — алгебры симметрий? Или это задача из динамики — и тогда к решению нужно привлекать ньютоновские дифференциальные уравнения?

Ответ таков: эта задача по определению относится к обеим названным областям. Причем это не пересечение областей, где мог бы находиться материал, общий для обеих, — они почти не пересекаются. Нет, это новая «область», охватывающая два традиционных раздела математики. Она как мост через реку, разделяющую две страны, связывает их, но не принадлежит ни одной. Но этот мост — не узкая полоса дороги: по размерам его можно сравнить с каждой из соединяемых стран. И, что еще важнее, используемые здесь методы не ограничиваются теми, что используются на прилежащих территориях. Фактически в моих исследованиях пригодились знания во всех областях математики, которые я когда-либо изучал. Так, курс по теории Галуа, который я слушал в Кембридже студентом, был посвящен решению (или, точнее, анализу причин, по которым мы не можем их решить) алгебраических уравнений пятой степени. В курсе по теории графов говорилось о сетях, т. е. о точках, соединенных линиями. Я не занимался динамическими системами, поскольку защищал докторскую по алгебре, но с годами познакомился с основными понятиями по этой теме — от статических состояний до хаоса. Итак, теория Галуа, теория графов, динамические системы: три отдельные области. По крайней мере я считал их таковыми до 2011 г., когда меня вдруг заинтересовал вопрос распознавания хаотической динамики в сети динамических систем, и тогда необходимым для исследования оказалось все то, что я узнал 45 лет назад на курсе по теории Галуа.

Итак, математика не похожа на политическую карту мира, где страны разделяются четкими границами и аккуратно окрашиваются каждая в свой цвет: розовый, зеленый или голубой. Она скорее напоминает естественный ландшафт, где никогда нельзя сказать наверняка, где заканчивается долина и начинаются предгорья, где лес переходит в лесостепь, кустарниковые заросли и настоящие степи, где озера вплавляют в окружающий ландшафт свои водяные зеркала, а реки связывают заснеженные горные склоны с далеким океаном. Но этот вечно меняющийся математический ландшафт состоит не из скал, воды и растений, а из идей, и соединяет все вместе не география, а логика. К тому же это динамичный ландшафт: он изменяется с появлением новых идей, с каждым новым открытием, с изобретением каждого нового метода. Важные концепции с множеством приложений подобны горным пикам, универсальные методики — широким рекам, несущим путешественников через плодородные равнины. Чем четче вырисовывается ландшафт, тем проще разглядеть на нем непокоренные еще вершины или неисследованные местности, которые часто воздвигают перед путником неожиданные и нежеланные препятствия. Со временем некоторые из этих пиков и препятствий становятся знаковыми. Это и есть великие проблемы математики.

Что делает математическую задачу великой? Интеллектуальная глубина в сочетании с простотой и элегантностью. Плюс к тому она должна быть *сложной*. Кто угодно может взобраться на холмик, но Эверест — совсем другое дело. Сформулировать великую задачу обычно нетрудно, хотя условия могут быть как элементарными, так и очень специальными и понятными только профессионалу. Если Великая теорема Ферма и проблема четырех красок без особых пояснений понятны всякому, кто знаком со школьной математикой, то, к примеру, гипотезу Ходжа или теорию Янга–Миллса даже сформулировать невозможно без привлечения глубоких концепций с переднего края науки (в конце концов, последняя имеет непосредственное отношение к квантовой теории поля). Тем не менее для специалиста в соответствующей области формулировки этих проблем звучат просто и естественно. Для их изложения не нужны многие страницы непонятного текста. И, наконец, существуют задачи, для детального понимания которых требуется уровень хотя бы университетского курса математики. Но более общий уровень понимания существа проблемы — откуда она взялась, почему важна, что можно было бы сделать, имея ее решение, — как правило, доступен любому интересующемуся, и именно это я попытаюсь вам объяснить. Правда, гипотеза Ходжа — крепкий орешек в этом отношении, поскольку она очень технична и очень абстрактна. Однако это одна из семи математических задач тысячелетия, за решение которых Институт Клэя предлагает приз в 1 млн долларов, и потому о ней непременно стоит рассказать.

Великие задачи несут в себе громадный творческий потенциал: они помогают создавать новую математику. В 1900 г. на Международном конгрессе математиков в Париже Давид Гильберт прочел лекцию, в которой перечислил 23 важнейшие математические проблемы. Он не включил в свой список Великую теорему Ферма, но упомянул ее во вступительном слове. Надо отметить, что, когда выдающийся математик перечисляет великие, по его мнению, проблемы, остальные математики относятся к этому очень серьезно. Понятно, что ни одна задача не оказалась бы в этом списке, не будь она важной и сложной.

Для человека естественно отвечать на вызов и преодолевать препятствия. С тех самых пор решение одной из гильбертовых проблем стало отличным способом завоевать себе математические «золотые шпоры». Многие из этих задач слишком специальны, чтобы включать их в эту книгу, другие представляют собой скорее программу, направление исследований, чем конкретные задачи, а некоторые мы рассмотрим позже по отдельности. Но сам список тоже заслуживает упоминания, и я включил его с кратким комментарием в примечания[1].

Именно это делает великие математические задачи великими. Проблема редко заключается в том, чтобы найти ответ. Математики очень четко представляют себе, какими должны быть ответы буквально всех великих задач, — или представляли, если на сегодняшний день решение уже известно. В самом деле, ожидаемый ответ часто заключен уже в формулировку вопроса. Гипотеза представляет собой правдоподобную догадку, предположение, основанное на совокупности данных. Как правило, хорошо изученные гипотезы со временем находят подтверждение, хотя так происходит не всегда. А в случае теоремы Ферма слово «теорема» употребляется (или, точнее, употреблялось) неверно — у теоремы обязательно должно быть доказательство, а его-то, пока не появился Уайлс, и не хватало.

Доказательство — вот то, чего требуют великие задачи и что делает их такими сложными. Любой человек, обладающий определенными знаниями, способен провести несколько вычислений, заметить явную закономерность и кратко сформулировать ее суть. Но математики требуют большего: они настаивают на полном, логически безупречном доказательстве. Или, если гипотеза не подтверждается, на столь же полном опровержении. Вообще же невозможно оценить всю чарующую привлекательность великой задачи, не понимая до конца жизненно важную роль доказательства в любом математическом предприятии. Обоснованное предположение может сделать кто угодно, трудно лишь доказать его истинность. Или ложность.

Концепция математического доказательства менялась с течением времени, причем требования к логике, как правило, становились все строже. Многочисленные высокоинтеллектуальные философские дискуссии о природе доказательства поднимали важные вопросы. Предлагались и внедрялись точные определения понятия «доказательство». Сегодня мы учим студентов, что доказательство начинается с набора некоторых явных допущений, известных как аксиомы. Аксиомы — это, так сказать, правила игры. В принципе возможны и другие аксиомы, но они относятся к другим играм. Первым такой подход предложил древнегреческий математик Евклид, но и сегодня он вполне применим. Доказательство на основе принятых аксиом представляет собой серию шагов, каждый из которых является логическим следствием либо аксиом, либо уже доказанных утверждений, либо того и другого. По существу, математика исследует логический лабиринт, перекрестками в котором служат утверждения, а проходами — достоверные умозаключения. Доказательство — путь через лабиринт, который начинается с аксиом. Утверждение, на котором он заканчивается, и есть то, что требовалось доказать.

Однако такое правильное и «причесанное» представление о доказательстве — еще не вся история и даже не самая главная ее часть. Это все равно что сказать: симфония — последовательность музыкальных нот, которая подчиняется законам гармонии. Определение верно, но где же творчество? Такое определение ничего не говорит нам не только о том, как искать доказательство, но и о том, как проверить его, когда оно предложено кем-то другим. Это определение ничего не говорит нам о том, какие места в лабиринте важнее других. Не говорит и о том, какие проходы в нем элегантны, а какие безобразны, какие значительны, а какие бесполезны. Это всего лишь формальное, механическое описание процесса, у которого немало и других аспектов, в частности человеческое измерение. Доказательства ищут люди, и математические исследования — отнюдь не воплощение пошаговой логики.

Формальный подход к определению доказательства может породить доказательства почти нечитаемые, поскольку основные усилия придется бросить на копание в мелочах и «расставление точек над логическими *i*», в то время как решающий вывод будет буквально бросаться в глаза. Поэтому практикующие математики спрямляют путь и оставляют за бортом все рутинные или очевидные шаги. На пропуски обычно указывают фразы вроде «несложно показать, что…» или «из стандартных расчетов следует, что…» Зато ни один математик не пройдет — по крайней мере сознательно — мимо логической трудности и не попытается сделать вид, что ее нет. Более того, компетентный математик постарается обратить особое внимание на слабые с точки зрения логики звенья цепочки рассуждений и потратит бóльшую часть времени и усилий на то, чтобы укрепить их и сделать достаточно надежными. Дело в том, что на практике доказательство — это математическая история с собственным сюжетом. У нее есть завязка, кульминация и развязка. В ней часто можно обнаружить боковые сюжетные ходы, которые вырастают из основного ствола, но ведут каждый к своему результату. Британский математик Кристофер Зиман однажды заметил, что любая теорема — это своего рода интеллектуальная точка покоя, где можно сделать остановку, перевести дыхание и ощутить некоторую определенность. Побочная сюжетная линия помогает свести концы с концами в основном сюжете. Доказательство напоминает литературный сюжет и в других отношениях: в них часто имеются один или несколько главных героев — конечно, это не люди, а идеи, — сложные взаимоотношения которых ведут к развязке и финалу.

Как явствует из формального определения, доказательство начинается с неких четких предположений, движется шаг за шагом от одного логического вывода к другому и заканчивается выводом о том, что вы, собственно, хотели доказать. Но доказательство — не просто список последовательных умозаключений, и логика в нем — не единственный критерий.

Доказательство — это рассказ, который выслушивают и разбирают по косточкам люди, посвятившие большую часть жизни искусству прочтения таких историй и поиска в них ошибок и противоречий. Основная цель этих людей — доказать, что автор доказательства *не прав*. Эти люди обладают поразительной способностью замечать слабые места и без устали долбить в них, пока вся конструкция не рухнет, подняв облако пыли. Вообще, если какой-нибудь математик заявляет, что ему удалось решить крупную проблему (одну из великих, например, или что-нибудь попроще, но тоже достойное), остальные математики не спешат кричать «Ура!» и открывать шампанское. Профессиональный инстинкт велит им прежде всего постараться опровергнуть предложенное доказательство.

Так или иначе, доказательство — это единственный надежный инструмент, при помощи которого математики могут убедиться в собственной правоте. Предвидя реакцию математического сообщества, исследователи тратят огромные усилия на проверку собственных выводов и поиск противоречий в них. Так проще. Если же история успешно выдерживает критический анализ коллег, сообщество вскоре приходит к выводу, что она верна, и в этот момент создатель доказательства получает заслуженные похвалы и награды. Во всяком случае, обычно бывает именно так, хотя непосредственным участникам событий это может видеться иначе. Когда ты вовлечен во что-то, то воспринимаешь все не так, как сторонний наблюдатель.

Как математики решают задачи? Этот вопрос почти не изучался. Современные образовательные исследования на базе когнитивистики в основном ограничиваются изучением образования от начальной до высшей школы. Есть исследования, посвященные преподаванию математики в вузах, но их не так уж много. Кроме того, есть большая разница между освоением и преподаванием математики и новыми исследованиями в этой области. Многие из нас умеют играть на каком-нибудь музыкальном инструменте, но мало кто способен сочинить симфонический концерт или хотя бы написать популярную песенку.

Когда речь заходит о творчестве на высочайшем уровне, почти все, что мы знаем — или думаем, что знаем, — мы получаем путем самоанализа. Мы просим математиков объяснить ход их мыслей и пытаемся выделить в этих описаниях общие принципы. Одной из первых серьезных попыток понять, как думают математики, можно считать книгу Жака Адамара «Исследование психологии процесса изобретения в области математики»[1], вышедшую в 1945 г. Адамар расспросил ведущих математиков и физиков своего времени и попросил описать, как они думают в процессе работы над сложной задачей. И тут выявилась важная и даже необходимая роль того, что за неимением лучшего термина следует назвать интуицией. Их мысли направляло нечто подсознательное. Самые плодотворные их идеи и озарения не приходили постепенно, в результате логической пошаговой проработки, а возникали неожиданно, и весь процесс развивался скачкообразно.

Одно из самых подробных описаний этого на первый взгляд нелогичного подхода к логическим вопросам дал французский математик Анри Пуанкаре — один из ведущих ученых конца XIX — начала XX в. Пуанкаре отметился едва ли не во всех областях математической науки, внес радикальные изменения во многие из них и основал несколько новых ее разделов. В последующих главах мы не раз будем возвращаться к его работам. Кроме того, Пуанкаре писал научно-популярные книги, и, возможно, именно огромный опыт и широта кругозора помогли ему глубже понять процесс собственного мышления. Во всяком случае, он был твердо убежден, что осознанная логика — лишь часть творческого процесса. Да, бывают моменты, когда без нее не обойтись: к примеру, без логики невозможно понять, в чем именно состоит проблема, как невозможно

---

[1] Адамар Ж. Исследование психологии процесса изобретения в области математики. — М.: Советское радио, 1970.

и проверить полученный ответ. Но в промежутке, считал Пуанкаре, его мозг нередко работал над задачей самостоятельно, ничего не сообщая хозяину, причем работал так, что хозяин был просто не в состоянии постичь его методы.

Его описание творческого процесса различает три ключевых этапа: подготовка, вынашивание и озарение. Подготовка представляет собой сознательные логические усилия, направленные на то, чтобы увидеть проблему, точно сформулировать ее и попробовать решить традиционными методами. Этот этап, когда подсознание получает задание и материал для работы, Пуанкаре считал очень важным. Вынашивание происходит, когда вы прекращаете думать о задаче, отвлекаетесь от нее и занимаетесь чем-то другим. А подсознание тем временем начинает перебирать и комбинировать идеи, часто довольно дикие, и продолжается это до тех пор, пока вдали не забрезжит свет. Если повезет, результатом станет озарение: подсознание даст вам сигнал, и в вашем мозгу как будто вспыхнет лампочка — возникнет готовый ответ.

Такое творчество подобно хождению по натянутому канату. С одной стороны, вы не можете решить сложную проблему, пока не познакомитесь как следует с областью, к которой она относится, а также с множеством других тем, которые могут пригодиться, а могут и не пригодиться в работе, просто на всякий случай. С другой стороны, если, изучая все нужные области математики, вы обратитесь к стандартному, уже много раз безрезультатно опробованному пути, то, возможно, уже не сумеете выбраться из наезженной колеи и ничего нового не откроете. Фокус в том, чтобы много знать и сознательно собирать свои знания воедино, работать над этим неделю за неделей… а затем отложить проблему в сторону. Тогда за дело возьмется интуитивная часть вашего сознания: она отсмотрит все идеи, повертит их так и эдак, оценит, где «холодно», а где «горячо», и сообщит вам, если что-нибудь найдет. Произойти это может в любой момент: Пуанкаре однажды понял, как нужно решать задачу, мучившую его несколько месяцев, выходя из автобуса.

Шриниваса Рамануджан, индийский математик-самоучка, создававший замечательные формулы, часто видел новые идеи во сне. А Архимед, согласно легенде, нашел способ определить содержание золота в сплаве, принимая ванну.

Пуанкаре особо указал, что без первоначального периода подготовки успеха не достичь. Подсознанию, настаивал он, необходимо дать как можно больше пищи для размышления, в противном случае удачные идеи, которые в конечном итоге могут привести к решению, просто не возникнут. Вдохновения без трудового пота не бывает. Кроме того, Пуанкаре наверняка знал — ведь об этом знает любой математик-исследователь, — что одного такого трехэтапного процесса редко бывает достаточно. Решение серьезной задачи, как правило, требует нескольких озарений. Этап вынашивания одной идеи может быть прерван вспомогательным процессом подготовки, вынашивания и озарения какой-то другой задачи, решение которой оказалось необходимым для работы над первой, основной идеей. Решение любой стоящей задачи, великой или не слишком, обычно включает в себя множество таких последовательностей, заключенных одна в другой, как замысловатые фракталы Бенуа Мандельброта. Вы решаете задачу, разбивая ее на подзадачи. Вы убеждаете себя, что если удастся решить эти подзадачи, то затем из полученных результатов можно будет собрать решение задачи в целом. Иногда они решаются, иногда приходится возвращаться к началу пути. Иногда подзадача сама рассыпается на несколько кусочков. Даже уследить за происходящим и удержать в голове общую картину порой очень и очень непросто.

Я назвал работу подсознания «интуицией». «Интуиция» — одно из удобных, но вводящих в заблуждение слов, таких как «инстинкт», которые широко используются, хотя и не имеют четкого значения. Подобными словами называют нечто непонятное, присутствие чего тем не менее отрицать невозможно. Математическая интуиция — это способность разума чувствовать форму и структуру и распознавать закономерности, которые мы не в состоянии уловить на сознательном уровне. Интуиция не обладает кристальной чистотой осознанной логики, зато способна привлечь наше внимание к вещам, которые мы никогда не стали бы рассматривать сознательно. Нейробиологи еще только начинают понимать, как человеческий мозг справляется с гораздо более простыми задачами. Понятно, однако, что интуиция, как бы она ни работала, существует благодаря структуре мозга и его взаимодействию с внешним миром.

Зачастую главное, чем помогает в работе интуиция, — она подсказывает, где у задачи слабые места, где к ней можно подступиться с максимальными шансами на успех. Математическое доказательство подобно сражению или, если вы предпочитаете менее воинственные сравнения, шахматной партии. Как только потенциально слабое место выявлено, исследователь бросает в бой (т. е. на его изучение) все свои возможности исследователя, весь математический аппарат, которым владеет. Как Архимед нуждался в точке опоры, чтобы перевернуть Землю, так и математик-исследователь нуждается в рычагах воздействия на задачу. Одна-единственная ключевая идея может раскрыть ее, сделать доступной для стандартных методов. Ну а после этого довести решение задачи до конца — дело техники.

Мой любимый пример рычагов такого рода — задачка, которая не имеет особого математического смысла, но помогает объяснить важный момент. Предположим, у вас есть шахматная доска из 64 клеток и набор костяшек домино, каждая из которых по размеру точно закрывает две соседние клетки доски. Очевидно, 32 костяшек достаточно, чтобы закрыть всю доску. Но теперь представьте, что из доски удалили две противоположных по диагонали угловых клетки, как показано на рис. 1. Можно ли закрыть оставшиеся 62 клетки при помощи 31 костяшки? Попробовав, вы поймете, что ничего не получается. С другой стороны, явных причин, по которым это задание можно было бы счесть невыполнимым, вроде бы тоже не видно. Но ровно до тех пор, пока вы не сообразите, что каждая костяшка домино, как их ни раскладывай, должна закрывать одну черную и одну белую клетку доски. Вот ваш рычаг, и теперь остается только применить его. Он подразумевает, что любая площадь, закрытая костяшками домино, содержит равное число черных и белых клеток. Но противоположные по диагонали клетки — одного цвета (в данном случае — белые), так что при их удалении возникает фигура, в которой черных клеток на две больше, чем белых. А никакую фигуру такого рода полностью закрыть костяшками невозможно. Наблюдение о том, что *любая* костяшка домино обязательно закрывает две клетки разного цвета, и есть слабое место этой головоломки. Поняв это, вы получаете точку, к которой можно приложить логический рычаг — и нажать. Если бы вы были средневековым бароном и осаждали замок, это стало бы для вас слабым местом замковой стены — местом, где следует сосредоточить огонь требушетов или начать делать подкоп.

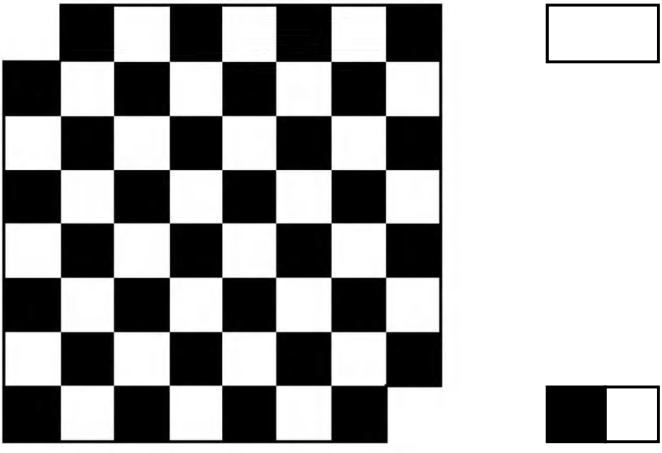


Рис. 1. Сможете ли вы закрыть обрезанную с двух углов доску костяшками домино, каждая из которых закрывает ровно две клетки (вверху справа)? Если вы раскрасите костяшку (внизу справа) и сосчитаете, сколько на доске черных и белых клеток, ответ станет очевиден

Однако в одном существенном моменте математические исследования отличаются от сражения. Любая территория, которую вам однажды удалось оккупировать, остается вашей навсегда, и после этого вы можете сосредоточить усилия на чем-то ином. Но доказанная теорема никуда не исчезает. И именно благодаря этому математики достигают прогресса в решении задачи, даже если дойти до конца им не удается. Однажды установленный факт становится доступен всем, и воспользоваться им может кто угодно и совершенно в любом контексте. Нередко отправной точкой новой атаки на древнюю как мир проблему становится незамеченное ранее сокровище, затерявшееся в целой куче разнообразных фактов. И это одна из причин, по которым любые новые математические расчеты ценны сами по себе, даже если польза от них не видна сразу. Это еще один кусок завоеванной территории, еще одно оружие в арсенале. Возможно, его время еще придет — но этого не случится, если его посчитают «бесполезным» и забудут или просто не дадут увидеть свет, потому что не поймут, какой в нем *смысл*.

2

Территория простых чисел

Проблема Гольдбаха

Некоторые великие задачи встречаются и в начальном курсе математики, хотя мы этого не замечаем. Вскоре после того, как ребенок осваивает умножение, он знакомится с концепцией простого числа. Известно, что некоторые числа могут быть получены при перемножении двух меньших чисел, к примеру: 6 = 2 × 3. Другие, такие как 5, невозможно разложить подобным образом на сомножители. Максимум, что можно сделать, это записать 5 = 1 × × 5, но в этом выражении нет двух *меньших* чисел. Числа, которые можно разбить на сомножители, называют составными, а те, что разложить невозможно, — простыми. Простые числа кажутся такой несложной темой! Если вы уже умеете перемножать натуральные числа, то способны разобраться и в том, что представляет собой простое число. Простые числа — первичные строительные кирпичики для всех натуральных чисел, и обнаружить их можно в самых разных разделах математики. Но в них есть тайна, и, на первый взгляд, они раскиданы среди положительных целых чисел почти случайным образом. Нет никаких сомнений: простые числа — настоящая загадка. Возможно, это естественное следствие их определения — ведь определяются они не через какое-либо присущее им свойство, а напротив — через свойство, которое у них отсутствует. С другой стороны, для математики это фундаментальное понятие, поэтому мы не можем просто так в ужасе поднять руки и сдаться. Нам необходимо с ними освоиться и каким-то образом вызнать их потаенные секреты.

Некоторые свойства простых чисел очевидны. За исключением самого маленького из них, двойки, все они нечетные. Сумма цифр простого числа, за исключением тройки, не может быть кратна трем. Они, за исключением пятерки, не могут заканчиваться на цифру 5. Если же число не подпадает под эти правила — и под несколько других, более тонких, — то невозможно посмотреть на него и сразу сказать, простое это число или нет. Да, существуют формулы для простых чисел, но это в значительной степени обман. Эти формулы не дают никакой полезной новой информации о простых числах; это просто хитрый способ зашифровать определение «простоты» в виде формулы. Простые числа — как люди: каждое из них — личность, и они не подчиняются общим правилам.

За тысячелетия математики сумели постепенно расширить свои знания о простых числах. Время от времени и сегодня решаются новые серьезные проблемы, с ними связанные. Однако многие вопросы по-прежнему остаются нерешенными. Некоторые из них фундаментальны и легко формулируются, другие понятны немногим. В этой главе говорится о том, что мы знаем и чего не знаем об этих раздражающих своей неприступностью, но все же фундаментальных числах. Начинается она с установления некоторых базовых понятий: в частности, концепции разложения на простые множители — как представить заданное число в виде произведения простых чисел. Даже этот знакомый процесс заводит нас на глубину сразу же, как только мы начинаем задавать вопросы о по-настоящему эффективных методах поиска простых множителей конкретного числа. Как ни удивительно, определить, является ли данное число простым, относительно несложно, но если число составное, то отыскать его простые множители часто намного труднее.

Разобравшись в основах, перейдем к самой известной из нерешенных задач, связанных с простыми числами, — к проблеме Гольдбаха, которой уже 250 лет. В последнее время в работе над ней достигнут колоссальный прогресс, но полностью она пока не решена. А несколько других задач представят нам примеры того, что еще предстоит сделать в этой важной, но трудно поддающейся исследованию области математики.

Простые числа и разложение на множители знакомы нам из школьного курса арифметики, однако большинство интересных свойств простых чисел на этом уровне не рассматривают и никаких доказательств не представляют. Тому есть веские причины: доказательства даже самых очевидных, на первый взгляд, свойств удивительно сложны. Вместо этого школьников учат некоторым простым методикам работы с простыми числами, акцентируя внимание на вычислениях, где цифры относительно невелики. В результате наши первые впечатления от встречи с простыми числами, как правило, обманчивы.

Древние греки были знакомы с некоторыми базовыми свойствами простых чисел и знали, как их доказать. Простые числа и сомножители — основная тема Книги VII евклидовых «Начал», классического труда, посвященного геометрии. В этой книге имеется, в частности, геометрическое представление арифметических действий — деления и умножения. Греки предпочитали работать не с числами как таковыми, а с длинами линий (отрезков), но их результаты несложно переформулировать на языке чисел. Так, Предложение 16 Книги VII доказывает, что при перемножении двух чисел результат не зависит от того, в каком порядке берутся эти числа. Иными словами, $ab = ba$, фундаментальный закон алгебры.

В школьной арифметике простые делители используют для поиска наибольшего общего делителя двух чисел. К примеру, чтобы найти наибольший общий делитель чисел 135 и 630, мы раскладываем их на простые множители:

$$135 = 3^3 \times 5; 630 = 2 \times 3^2 \times 5 \times 7.$$

Затем берем все простые числа, которые присутствуют в обоих разложениях, в наибольшей общей степени; получаем $3^2 \times 5$. Перемножаем, получаем 45. Это и есть наибольший общий делитель. Из этой процедуры создается впечатление, что без разложения на простые множители невозможно найти наибольший общий делитель. На самом деле с точки зрения логики все наоборот. Предложение 2 Книги VII «Начал» представляет метод поиска наибольшего общего делителя двух натуральных чисел без разложения их на простые множители. Метод состоит в последовательном вычитании меньшего числа из большего, а затем остатка из меньшего числа и т. д. до тех пор, пока есть остаток. Для тех же чисел 135 и 630 — это достаточно типичный случай для небольших чисел — процесс выглядит так. Вычитаем 135 из 630 столько раз, сколько сможем:

$630 - 135 = 495;$
$495 - 135 = 360;$
$360 - 135 = 225;$
$225 - 135 = 90.$

Поскольку $90 < 135$, переходим к той же процедуре с участием чисел 90 и 135:

$135 - 90 = 45.$

Поскольку $45 < 90$, продолжаем то же с числами 45 и 90:

$90 - 45 = 45;$
$45 - 45 = 0.$

Таким образом, наибольший общий делитель чисел 135 и 630 равен 45.

Эта процедура работает потому, что на каждой стадии происходит замена первоначальной пары чисел более простой парой (одно из чисел уменьшается), которая тем не менее имеет тот же наибольший общий делитель. В конце концов, одно из чисел делится на второе нацело, без остатка, и процесс поиска на этом завершается. В наше время подробное описание вычислительного метода, при помощи которого можно гарантированно найти ответ той или иной задачи, называют алгоритмом. Поэтому и процедура из «Начал» Евклида известна сегодня как евклидов алгоритм. Логически эта процедура первична по отношению к процедуре разложения на простые множители. В самом деле, Евклид использует ее для доказательства основных свойств простых делителей. В современных университетских курсах математики алгоритм Евклида используется с той же целью.

Описанная процедура целиком опирается на евклидово Предложение 30 и была бы невозможна без него. В современных терминах речь в нем идет о том, что если произведение двух чисел — то, что мы получаем при их перемножении — делится на некое простое число, то на это же число должен делиться один из сомножителей. Предложение 32 заключается в том, что любое число либо само является простым, либо имеет простой делитель. Объединив оба утверждения, несложно сделать вывод, что любое число есть результат перемножения простых множителей и что их набор единственный, если не брать во внимание порядок записи. К примеру,

$$60 = 2 \times 2 \times 3 \times 5 = 2 \times 3 \times 2 \times 5 = 5 \times 3 \times 2 \times 2$$

и т. д., но единственный реальный способ получить 60 состоит в том, чтобы взять множители из первого разложения и переставить их местами. Не существует, к примеру, разложения, в котором 60 = 7 × *что-нибудь*. Существование какого-нибудь разложения следует из Предложения 32. Если число простое — стоп. Если нет, находим простой делитель, делим на него, получая меньшее число, и повторяем процедуру. Уникальность набора делителей следует из Предложения 30. Так, если бы разложение 60 = 7 × *что-нибудь* существовало, то одно из чисел

2, 3 и 5 должно было бы тоже делиться на 7, но этого не происходит.

Здесь я должен прояснить один небольшой, но важный момент: исключительный статус числа 1. Согласно приведенному выше определению, оно простое: если мы попытаемся разбить его на множители, максимум, что мы получим, будет 1 = 1 × 1, где нет меньших чисел. Однако позже, с развитием теории, такая интерпретация вызывает проблемы, поэтому в последние век-два математики добавили в определение простого числа дополнительное ограничение. Число 1 настолько отличается от всех остальных чисел, что его следует рассматривать как исключение, — это не простое число, но и не составное. Это третья разновидность числа — единица. Одна из причин, по которым мы называем 1 особым случаем, а не относим ее к настоящим простым числам, заключается в том, что если мы согласимся с простотой единицы, то единственность набора множителей нарушится. Вообще-то 1 × 1 = 1 — уже нарушение, а уж 1 × 1 × 1 × 1 × 1 × 1 × 1 × 1 = 1 ни в какие ворота не лезет. Можно было бы изменить определение единственности и сказать «единственный, без учета дополнительных единичных множителей», но это был бы всего лишь другой способ признать, что 1 — число особое.

Много позже, в Предложении 20 Книги IX, Евклид доказывает еще один ключевой факт: «Простых чисел существует больше, чем их насчитывается в любом множестве простых чисел». Иными словами, множество простых чисел бесконечно. Это чудесная теорема и изящное доказательство, но ее появление вызвало множество проблем. Если простые числа уходят в бесконечность, но, судя по всему, расположены без всякой системы, то как можно сказать, на что они похожи?

Мы вынуждены обратиться к этому вопросу потому, что не можем оставить в стороне простые числа — очень существенную деталь математического ландшафта. Особенно часто они встречаются (и особенно полезны) в теории чисел — раз-деле математики, изучающем свойства целых чисел. Звучит, может быть, достаточно элементарно, но на самом деле теория чисел — один из самых глубоких и сложных разделов математики. Позже мы увидим тому множество свидетельств. В 1801 г. Гаусс, ведущий специалист того времени по теории чисел (а также, по мнению некоторых ученых, один из ведущих математиков всех времен, а может быть, и величайший из них), написал продвинутый учебник по этой теории — «Арифметические исследования» (*Disquisitiones Arithmeticae*). В нем среди множества сложных тем Гаусс указал, что не следует терять из виду два весьма фундаментальных вопроса: «Известно, что задача отличения простых чисел от составных и разложения последних на простые множители является одной из важнейших и полезнейших в арифметике».

В школе, как правило, учат ровно одному способу поиска простых делителей числа. Заключается он в том, чтобы пробовать по очереди все потенциальные делители, пока не найдется такой, на который число разделится нацело. Если вы не нашли ни одного делителя к тому моменту, как добрались до корня квадратного из первоначального числа — точнее, до наибольшего целого числа, меньшего или равного этому корню, — то число это простое. В противном случае вы найдете множитель, разделите на него и продолжите с новым числом с того же места. Эффективнее всего пробовать только простые делители, но для этого необходим список простых чисел. Поиск останавливается на корне квадратном из числа, потому что наименьший делитель любого составного числа не превосходит корень квадратный из этого числа. Однако для больших чисел эта процедура безнадежно неэффективна. К примеру, если взять число

1 080 813 321 843 836 712 253,

то на простые множители оно раскладывается следующим образом:

13 929 010 429 × 77 594 408 257,

и, чтобы добраться до меньшего из двух множителей, вам придется опробовать каждое из первых 624 401 249 простых чисел. Конечно, при помощи компьютера это несложно сделать, но если взять для начала число из 100 цифр, которое — так уж случилось — раскладывается на два множителя по 50 цифр в каждом, то систематический перебор последовательных простых чисел продлится до конца Вселенной и вряд ли успеет дать результат.

Нет, вообще-то современные компьютеры, как правило, умеют раскладывать числа из 100 цифр на простые множители. Моему компу требуется меньше секунды, чтобы найти простые множители числа $10^{99} + 1$ (выглядит это число как 1000 … 001 с 98 нулями). Это число — результат перемножения 13 простых чисел (одно из них повторяется дважды), наименьшее из которых — 7, а наибольшее — 141 122 524 877 886 182 282 233 539 317 796 144 938 305 111 168 717.

Однако если я попрошу компьютер разложить на множители число $10^{199} + 1$, в котором 200 цифр, то жужжать он будет долго, но результата так и не выдаст. Хотя, конечно, даже разложение числа из 100 цифр производит сильное впечатление. В чем тут секрет? В более эффективном по сравнению с последовательным перебором потенциальных простых делителей алгоритме поиска.

Мы сегодня знаем о первой из названных Гауссом задач (проверка числа на простоту) гораздо больше, чем знал он сам, и гораздо меньше, чем хотелось бы, о второй (разложение на простые множители). Здравый смысл говорит о том, что проверка на простоту намного проще разложения на простые множители. Как правило, это удивляет нематематиков, — ведь в школе учат проверять число на простоту тем же методом, что и искать его простые множители: перебором всех возможных делителей. Но, оказывается, существуют хитрые способы доказать простоту числа и без этого. Эти же методы позволяют доказать, что число составное, без нахождения каких бы то ни было его делителей. Достаточно показать, что это число не проходит тест на простоту.

Прапрадедушкой всех современных тестов на простоту может считаться теорема Ферма (чтобы не путать со знаменитой Великой теоремой, о которой речь пойдет в главе 7, ее иногда называют Малой теоремой Ферма). Эта теорема основана на модулярной арифметике, которую иногда называют еще «часовой арифметикой», поскольку числа в ней спирально накладываются друг на друга, как время на циферблате часов. Выберем число — для 12-часовых аналоговых часов это число 12 — и назовем его модулем. Теперь в любых арифметических вычислениях с неотрицательными целыми числами мы договоримся заменять любое число, кратное 12, нулем. К примеру, $5 \times 5 = 25$, но 24 — это дважды 12, поэтому вычтем из результата 24. Получим $5 \times 5 = 1$ по модулю 12. Модулярная арифметика очень красива, поскольку почти все обычные арифметические законы в ней тоже работают. Основная разница заключается в том, что мы не всегда можем разделить одно число на другое, даже если это не нуль. Модулярная арифметика полезна также тем, что обеспечивает удобный и аккуратный способ разбираться с вопросами делимости: какие числа делятся на те или иные модули без остатка и чему равен остаток, если это не так. Модулярную арифметику предложил Гаусс в «Арифметических исследованиях», и сегодня она широко используется не только в математике, но и в информатике, физике, инженерном деле.

Малая теорема Ферма утверждает, что если взять простой модуль p и любое число a, не кратное p, то степень $(p - 1)$ числа a будет равна 1 по модулю p. Пусть, к примеру, $p = 17$ и $a = 3$. Тогда теорема предсказывает, что остаток от деления 3^{16} на 17 будет равен 1. Проверим:

$$3^{16} = 43\,046\,721 = 2\,532\,160 \times 17 + 1.$$

Ни один человек, находящийся в своем уме, не захочет проводить подобные расчеты для, скажем, 100-значных простых чисел. К счастью, существует хитрый и быстрый способ сделать это. Смысл в том, что ответ не равен единице, если модуль, с которого мы начали, является составным числом. Так что теорема Ферма — надежная основа для эффективного теста, который обеспечивает необходимое условие простоты числа.

К несчастью, одного этого теста недостаточно. Известно, что его проходят и многие составные числа, известные как числа Кармайкла. Самое маленькое из них 561, и в 2003 г. Ред Элфорд, Эндрю Гранвиль и Карл Померанс доказали, к всеобщему изумлению, что таких чисел бесконечно много. Изумление математического сообщества вызвал тот факт, что авторам удалось найти доказательство; сам по себе результат особого удивления не вызвал. Фактически было доказано, что для каждого числа x существует по крайней мере $x^{2/7}$ чисел Кармайкла, меньших или равных x, если x достаточно велико.

Однако более сложные варианты теоремы Ферма действительно можно превратить в тесты на простоту, такие как опубликованный в 1976 г. Гэри Миллером. К несчастью, доказательство достоверности теста Миллера опирается на одну из нерешенных великих математических задач — обобщенную гипотезу Римана (глава 9). В 1980 г. Майкл Рабин превратил тест Миллера в вероятностный, т. е. такой, который может иногда давать неверный ответ. Исключения, если они существуют, встречаются очень редко, но тем не менее доказать, что их нет, невозможно.

Наиболее эффективным детерминированным (т. е. дающим гарантированный результат) тестом на сегодняшний день является тест Адлемана–Померанса–Румели, названный в честь своих создателей — Леонарда Адлемана, Карла Померанса и Роберта Румели. В нем используются концепции теории чисел, куда более сложные, чем теорема Ферма, но примерно того же характера.

Я до сих пор помню письмо одного математика-любителя, предложившего вариант испытания делением. Давайте пробовать все возможные делители, предлагал этот энтузиаст, но начинать с корня квадратного из числа и двигаться, наоборот, вниз. Иногда этот метод действительно позволяет быстрее получить результат, чем при проверке делителей в обычном порядке, но с ростом чисел он, естественно, встречается с теми же проблемами, что и обычный метод. Если применить предложенный вариант к приведенному выше примеру, 22-значному числу 1 080 913 321 843 836 712 253, то квадратный корень из него равен примерно 32 875 725 419. Вам придется перепробовать 794 582 971 простой делитель, прежде чем вы доберетесь до нужного. Это *хуже,* чем искать его обычным путем.

В 1956 г. знаменитый логик Курт Гедель в письме к Джону фон Нейману почти буквально повторил мольбу Гаусса. Он спрашивал, можно ли улучшить метод пробного деления, и если можно, то насколько. Фон Нейман не стал заниматься этим вопросом, но позже другие математики ответили Геделю, открыв практические методы нахождения простых чисел длиной до 100 знаков, а иногда даже больше. Эти методы, самый известный из которых называется методом квадратичного решета, появились около 1980 г. Однако почти все они либо вероятностны, либо неэффективны в следующем смысле.

Как увеличивается компьютерное время, необходимое для вычислений, с ростом объема исходных данных? При тестировании на простоту исходные данные — это не само число, а число знаков в нем. Ключевое различие в этом случае проводится между двумя группами алгоритмов — алгоритмами, принадлежащими и не принадлежащими к классу P. Если время работы алгоритма растет как некая фиксированная степень от размера исходных данных, то алгоритм принадлежит к классу P; в противном случае — не принадлежит. Грубо говоря, алгоритмы класса P полезны, тогда как те, что не принадлежат к этому классу, непрактичны. Существует, однако, промежуточная полоса своеобразной ничьей земли, где в ход идут другие соображения. Класс P получил название от понятия «полиномиальное время» — именно так замысловато математики говорят о постоянных степенях. Мы еще вернемся к теме эффективных алгоритмов позже, в главе 11.

По стандартам класса P метод пробного деления работает из рук вон плохо. На школьном уровне, где для проверки предлагаются двух- или трехзначные числа, с ним все в порядке, но при работе со 100-значными числами он абсолютно безнадежен. В общем, пробное деление никак не укладывается в P-класс. Если быть точным, то время выполнения этого алгоритма для любого n-значного числа приблизительно равняется $10^{n/2}$, а эта величина растет быстрее, чем любая фиксированная степень n. С таким типом роста, известным как экспоненциальный, по-настоящему *трудно* иметь дело, это страшный сон любого, кто занимается вычислениями.

До 1980-х гг. у всех известных алгоритмов проверки на простоту, за исключением вероятностных или тех, надежность которых оставалась недоказанной, время вычислений росло экспоненциально. Однако в 1983 г. был найден алгоритм, очень соблазнительно лежащий на ничьей земле вблизи P-территории: это уже упоминавшийся тест Адлемана–Померанса–Румели. Его улучшенная версия, разработанная Генри Коэном и Хендриком Ленстрой, имела время вычисления n в степени $\log \log n$, где $\log$ — обозначение логарифма. Технически $\log \log n$ может быть сколь угодно большим, поэтому данный алгоритм не относится к P-классу. Однако это не мешает ему быть пригодным к практическому использованию: если n — гуголплекс, т. е. 1 с 10^{100} нулями, то $\log \log n$ равен примерно 230. Старая шутка гласит: «Доказано, что $\log \log n$ стремится к бесконечности, но никто никогда не видел, как он это делает».

Первый тест на простоту, принадлежащий к P-классу, открыли в 2002 г. Маниндра Агравал и его студенты-дипломники Нирадж Каял и Нитин Саксена. В Примечаниях можно прочитать об этом немного подробнее[2]. Они придумали алгоритм и доказали, что время его выполнения растет пропорционально не более чем n^{12}; очень скоро эта величина была уменьшена до $n^{7,5}$. Однако, несмотря на то что их алгоритм относится к P-классу и, соответственно, считается «эффективным», его преимущества не проявляются до тех пор, пока n не становится очень и очень большим. По идее этот алгоритм должен побить тест Адлемана–Померанса–Румели, когда *число знаков* в n приблизится к 10^{1000}. Но такое большое число невозможно разместить не только в память компьютера, но и вообще в известной Вселенной. Зато теперь мы *точно знаем,* что алгоритмы P-класса для проверки простоты числа существуют. Ясно, что поиск лучших алгоритмов в этой категории — дело ст*о*ящее. Ленстра и Померанс снизили степень с 7,5 до 6. Если еще некоторые предположения о свойствах простых чисел подтвердятся, степень можно будет снизить до 3, что приблизит нас к практическому применению подобных алгоритмов.

Но самое интересное в алгоритме Агравала–Каяла–Саксены — не результат, а метод. Он прост — по крайней мере для математиков — и отличается новизной. В основе его лежит вариант теоремы Ферма, но, вместо того чтобы работать с числами, команда Агравала использовала многочлены. Многочлен, или полином, — это комбинация степеней переменной x, такая, к примеру, как $5x^3 + 4x - 1$. Многочлены можно складывать, вычитать и перемножать, и обычные алгебраические законы на них тоже распространяются. В главе 3 мы поговорим о многочленах подробнее.

По-настоящему великолепная идея: расширить пространство дискурса и перенести проблему в новую область. Это тот самый случай, когда идея проста настолько, что нужно быть гением, чтобы разглядеть ее. Первый намек на нее проскользнул в статье Агравала и его научного консультанта Сомената Бисваса: авторы предложили вероятностный тест на простоту, основанный на аналоге теоремы Ферма в мире полиномов. Агравал был убежден, что вероятностный компонент этого метода может быть устранен. В 2001 г. его студенты пришли к нему с очень важным техническим замечанием. Начав в нем разбираться, команда углубилась в дебри теории чисел, но постепенно, со временем, все замечания удалось свести к единственному препятствию — вопросу существования простого числа p, такого, чтобы число $p - 1$ имело бы достаточно большой простой делитель. Несколько консультаций с коллегами и поиск в Интернете помогли обнаружить теорему, которую Этьен Фуври доказал в 1985 г. при помощи сложных формальных методов. Именно этого команде Агравала недоставало, чтобы доказать работоспособность алгоритма, и последняя деталь головоломки точно встала на место.

В те времена, когда теория чисел пребывала в своей башне из слоновой кости, вся эта история прошла бы незамеченной и никак не повлияла бы на жизнь остального мира. Но в последние 20 лет простые числа приобрели огромный вес в криптографии — науке о шифрах. Шифры важны не только для военных, у коммерческих компаний тоже хватает секретов. Сегодня, в век Интернета, секреты есть у каждого из нас: мы не хотим, чтобы преступники получили доступ к нашим банковским счетам и номерам кредитных карт. Мало того, все чаще в преступных целях используются и другие личные данные, так что хотелось бы уберечь их все, вплоть до клички домашней кошки. Но Интернет невероятно удобен при оплате счетов, страховании машин и заказе всего, что необходимо для поездки на отдых, и всем нам приходится мириться с риском того, что ценная частная информация попадет не в те руки.

Производители компьютеров и интернет-провайдеры пытаются снизить этот риск, предлагая пользователям различные системы шифрования. Надо сказать, что внедрение компьютеров изменило как саму криптографию, так и криптоанализ — искусство взлома шифров. В настоящее время разработано множество новых шифров. Один из самых известных шифров, который в 1978 г. придумали Рональд Ривест, Ади Шамир и Леонард Адлеман, основан на использовании простых чисел. Больших простых чисел, примерно 100-значных. Система

Ривеста–Шамира–Адлемана (известная как RSA) используется во многих компьютерных операционных системах, встроена в основные протоколы безопасного интернет-соединения, ею широко пользуются правительства, корпорации и университеты. Конечно, не каждое новое открытие, имеющее отношение к простым числам, может повлиять на безопасность вашего банковского счета, но это добавляет теме интереса. Как только удается выяснить что-то новое, что помогает связать простые числа и компьютерные вычисления, это привлекает повышенное внимание. Так случилось и с тестом Агравала–Каяла–Саксены, хотя при всей своей математической элегантности и важности непосредственного практического значения он не имеет.

Тем не менее он позволил немного под другим углом рассмотреть общий вопрос криптографии по Ривесту–Шамиру–Адлеману, и результат вызывает некоторые опасения. До сих пор не существует ни одного алгоритма P-класса для решения второй из названных Гауссом задач — разложения на простые множители. Большинство специалистов сходятся во мнении, что такого алгоритма не существует, но в последнее время их уверенность несколько поколебалась. Поскольку где-то за кулисами, совсем рядом, могут скрываться и другие открытия, подобные тесту Агравала–Каяла–Саксены и основанные на таких же простых идеях, как полиномиальная версия теоремы Ферма (и не важно, что пока о них никто даже не подозревает), может оказаться, что системы шифрования, основанные на разложении числа на простые множители, не настолько надежны, как нам хочется верить. Так что пока не стоит раскрывать в Интернете кличку вашей кошки!

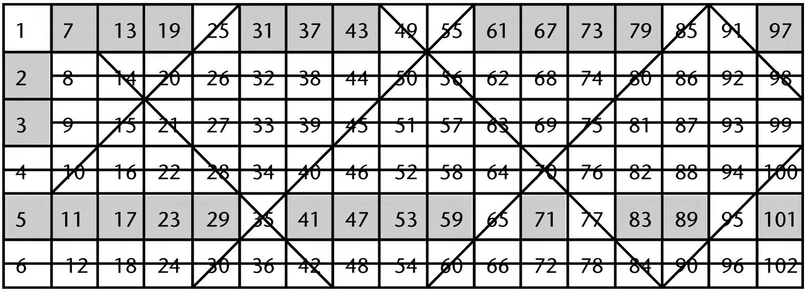
Даже элементарная математика простых чисел ведет к выдвижению более сложных концепций. Евклид доказал, что простые числа уходят в бесконечность, так что невозможно просто перечислить их все и успокоиться. Мы не можем также дать простую и практичную алгебраическую формулу для вычисления всех простых чисел подряд, примерно так, как по формуле x^2 вычисляются квадраты чисел. (Простые формулы существуют, но они «мошенничают», встраивая в формулу сами простые числа под разными личинами, и в результате не сообщают нам ничего нового[3].) Пытаясь познать природу этих неуловимых и странных чисел, мы экспериментируем, ищем в них признаки структурированности и пытаемся доказать, что найденные нами закономерности присутствуют во всех простых числах, какими бы большими они ни были. Можно, к примеру, задаться вопросом о том, как простые числа распределены среди всех целых чисел. Таблицы простых чисел позволяют предположить, что чем дальше, тем таких чисел становится меньше. В табл. 1 показано, сколько простых чисел содержится в разных диапазонах на 1000 последовательных целых чисел.

Таблица 1. Количество простых чисел в последовательных интервалах по 1000 чисел

Интервал	**Количество простых чисел**
1–1000	168
1001–2000	135
2001–3000	127
3001–4000	119
4001–5000	118
5001–6000	114
6001–7000	117
7001–8000	106
8001–9000	110
9001–10 000	111

Числа во второй колонке по большей части уменьшаются сверху вниз, хотя иногда ненадолго изменяют свое поведение: к примеру, после 114 мы видим 117. Это симптом нерегулярности простых чисел, но в целом общая тенденция прослеживается достаточно четко: чем больше числа, тем реже среди них встречаются простые. За объяснением не нужно далеко ходить: чем больше становится число, тем больше у него потенциальных делителей. А простые числа должны избегать каких бы то ни было делителей. Это напоминает ловлю составных (непростых) чисел рыболовной сетью: чем гуще становится сеть, тем меньшему числу простых чисел удается сквозь нее проскользнуть.

У этой «сети» есть даже название: решето Эратосфена. Эратосфен Киренский — древнегреческий математик, живший около 276–194 гг. до н. э. Он также был атлетом, интересовался поэзией, географией, астрономией и музыкой. Эратосфен первым сумел разумным образом оценить размеры Земли, обратив внимание на положение солнца в полдень в двух разных местах — Александрии и Сиене (современный Асуан). В Сиене солнце в полдень стояло точно над головой, а в Александрии отстояло от вертикали примерно на 7°. Поскольку угол в 7° составляет одну пятидесятую часть круга, то и окружность Земли должна в 50 раз превосходить расстояние от Александрии до Сиены. Эратосфен не мог непосредственно измерить это расстояние, поэтому он спросил у караванщиков, сколько времени занимает путешествие на верблюдах из одного города в другой, и оценил, сколько в среднем проходят верблюды за день. Результат своих расчетов он привел в тогдашних единицах расстояния — *стадиях,* но мы не знаем, чему равнялась стадия. Историки сходятся во мнении, что оценка Эратосфена оказалась достаточно точной.



1	7	13	19	~~25~~	31	37	43	~~49~~	~~55~~	61	67	73	79	~~85~~	~~91~~	97
2	~~8~~	~~14~~	~~20~~	~~26~~	~~32~~	~~38~~	~~44~~	~~50~~	~~56~~	~~62~~	~~68~~	~~74~~	~~80~~	~~86~~	~~92~~	~~98~~
3	~~9~~	~~15~~	~~21~~	~~27~~	~~33~~	~~39~~	~~45~~	~~51~~	~~57~~	~~63~~	~~69~~	~~75~~	~~81~~	~~87~~	~~93~~	~~99~~
~~4~~	~~10~~	~~16~~	~~22~~	~~28~~	~~34~~	~~40~~	~~46~~	~~52~~	~~58~~	~~64~~	~~70~~	~~76~~	~~82~~	~~88~~	~~94~~	~~100~~
5	11	17	23	29	~~35~~	41	47	53	59	~~65~~	71	~~77~~	83	89	~~95~~	101
~~6~~	~~12~~	~~18~~	~~24~~	~~30~~	~~36~~	~~42~~	~~48~~	~~54~~	~~60~~	~~66~~	~~72~~	~~78~~	~~84~~	~~90~~	~~96~~	~~102~~

Рис. 2. Решето Эратосфена

Решето Эратосфена представляет собой алгоритм поиска всех простых чисел путем последовательного исключения из числового ряда чисел, кратных уже известным простым. Рисунок 2 иллюстрирует этот метод на числах от 1 до 102, организованных так, чтобы процесс исключения кратных чисел был хорошо виден. Чтобы посмотреть, как все происходит, я советую вам составить эту или подобную ей схему самостоятельно, с нуля. Для начала начертите табличку и заполните ее числами, ничего не закрашивая и не перечеркивая. Затем потихоньку начинайте вычеркивать. Исключите 1, потому что это единица. Следующее число — 2, значит, оно простое. Вычеркните все числа, кратные 2: это те, что лежат на горизонталях, начинающихся с чисел 4, 6 и 8. Следующее невычеркнутое число — 3, следовательно, оно простое. Вычеркните все числа, кратные 3: это горизонтальный ряд, начинающийся с 6 (уже вычеркнут) и с 9. Следующее невычеркнутое число — 5, оно простое. Вычеркиваем все числа, кратные 5: они находятся на диагональных линиях, идущих слева снизу вверх направо и начинающихся на 10, 30, 60 и 90. Следующее невычеркнутое число — 7, оно простое. Вычеркиваем все числа, кратные 7: это диагонали, проходящие сверху слева вниз направо и начинающиеся на 14, 49 и 91. Затем 11 — оно не вычеркнуто, и это простое число. Первое число, кратное 11 и до сих пор не вычеркнутое (т. е. не имеющее меньших делителей) — 121, — находится за пределами нашей таблички. Процесс окончен. Оставшиеся числа в серых ячейках и есть искомые простые числа.

Решето Эратосфена — не просто историческая диковинка, это и сегодня один из наиболее эффективных методов составления длинных списков простых чисел. А родственные ему методы позволили достичь значительного прогресса в решении самой знаменитой, наверное, из великих нерешенных проблем, имеющих отношение к простым числам: проблемы Гольдбаха. Немецкий математик-любитель Кристиан Гольдбах переписывался со многими знаменитостями своего времени. В 1742 г. в письме к Леонарду Эйлеру он изложил несколько любопытных гипотез, связанных с простыми числами. Позже историки заметили, что Рене Декарт ранее писал примерно то же самое. Первое из утверждений Гольдбаха звучало так: «Всякое целое число, которое можно представить как сумму двух простых, можно записать также как сумму произвольного числа простых, пока все слагаемые не станут единицами». Второе утверждение, добавленное уже на полях письма, гласило: «Всякое целое число больше двух можно представить как сумму трех простых». Сегодняшнее определение простого числа предполагает очевидные исключения из обоих утверждений. Так, 4 не есть сумма трех простых, поскольку наименьшее простое число — 2, и сумма трех простых не может быть меньше 6. Однако во времена Гольдбаха число 1 считалось простым. Разумеется, его утверждения можно переформулировать в соответствии с современными представлениями.

В ответном письме Эйлер припомнил предыдущий разговор с Гольдбахом, когда тот указал, что первое его заявление является следствием более простой, третьей гипотезы: «Всякое четное целое есть сумма двух простых». С учетом общепринятого представления о 1 как о простом числе из этого утверждения прямо следует вторая гипотеза, поскольку любое число можно выразить как $n + 1$ или $n + 2$, где n — четное. Если n есть сумма двух простых, то исходное число есть сумма трех простых. Мнение Эйлера о третьем заявлении было однозначным: «Я считаю, что это, несомненно, верная теорема, хотя и не могу ее доказать». Собственно, на сегодняшний день статус этой гипотезы практически не изменился.

Современный подход, при котором 1 — не целое число, разбивает гипотезу Гольдбаха на две части. Вариант для четных чисел (так называемая бинарная проблема Гольдбаха) гласит: любое четное целое число больше двух можно представить в виде суммы двух простых чисел.

А вот вариант для нечетных (известный как тернарная проблема Гольдбаха): любое нечетное число больше 5 можно представить в виде суммы трех простых чисел.

Из бинарной гипотезы автоматически следует тернарная, но не наоборот[4]. Есть смысл рассматривать эти гипотезы по отдельности, поскольку мы до сих пор не знаем точно, верна ли хоть одна из них. Но, похоже, тернарная проблема немного проще, в том смысле что продвинуться в этом направлении удалось заметно дальше.

Бинарную гипотезу Гольдбаха для малых чисел можно подтвердить несложными вычислениями:

$4 = 2 + 2;$
$6 = 3 + 3;$
$8 = 5 + 3;$
$10 = 7 + 3 = 5 + 5;$
$12 = 7 + 5;$
$14 = 11 + 3 = 7 + 7;$
$16 = 13 + 3 = 11 + 5;$
$18 = 13 + 5 = 11 + 7;$
$20 = 17 + 3 = 13 + 7.$

Несложно продолжить ряд примеров вручную, скажем, до 1000 или около того, а можно и дальше, если хватит терпения. К примеру, $1000 = 3 + 997$, а $1\,000\,000 = 17 + 999\,983$. В 1938 г. Нильс Пиппинг проверил бинарную гипотезу Гольдбаха для всех четных чисел вплоть до 100 000.

При этом выявилась общая тенденция: чем больше само число, тем больше способов представить его в виде суммы простых. Это отвечает здравому смыслу. Если вы возьмете большое четное число и начнете вычитать из него по очереди простые числа, с какой вероятностью *все* результаты этих действий окажутся составными? Достаточно в списке разностей появиться хотя бы одному простому числу, — и можно считать, что гипотеза для исходного числа подтверждена. Обратившись к статистическим свойствам простых чисел, можно оценить вероятность такого исхода. В 1923 г. аналитики Харольд Харди и Джон Литлвуд проделали такую операцию и вывели правдоподобную, но нестрогую формулу для числа способов представления заданного четного n в виде суммы двух простых чисел: это число приблизительно равно $n/[2\ (\log n)^2]$. Это число увеличивается с ростом n и, кроме того, хорошо согласуется с числовыми данными. Но даже если математикам удалось бы сделать эту формулу точной, невозможно было бы исключить возможность того, что из нее существуют очень редкие, но все же исключения, так что формула не слишком помогает.

Основное препятствие, мешающее доказать гипотезу Гольдбаха, заключается в том, что она сочетает в себе две очень разные характеристики. Простые числа определяются через умножение, а в самой гипотезе речь идет о сложении. Поэтому необычайно трудно соотнести желаемый вывод с каким бы то ни было разумным свойством простых чисел. Такое впечатление, что рычаг просто некуда вставить. Должно быть, эти слова звучали настоящей музыкой в ушах владельцев издательства Faber & Faber, когда в 2000 г. они пообещали премию в 1 000 000 долларов за доказательство гипотезы. Сделано это было ради продвижения романа Апостолоса Доксиадиса «Дядя Петрос и проблема Гольдбаха»[1]. Сроки поджимали: решение необходимо было представить до апреля 2002 г. Премия эта так никому и не досталась, что едва ли удивительно, если учесть, что проблема Гольдбаха остается нерешенной уже более 250 лет.

Гипотезу Гольдбаха часто формулируют иначе — как вопрос о сложении множеств целых чисел. Бинарная проблема Гольдбаха — простейший пример такого подхода, поскольку при этом мы складываем всего лишь два множества. Для этого нужно взять любое число из первого множества, добавить к нему любое число из второго и составить из всех таких сумм свое, третье множество. Так, сумма множеств $\{1, 2, 3\}$ и $\{4, 5\}$

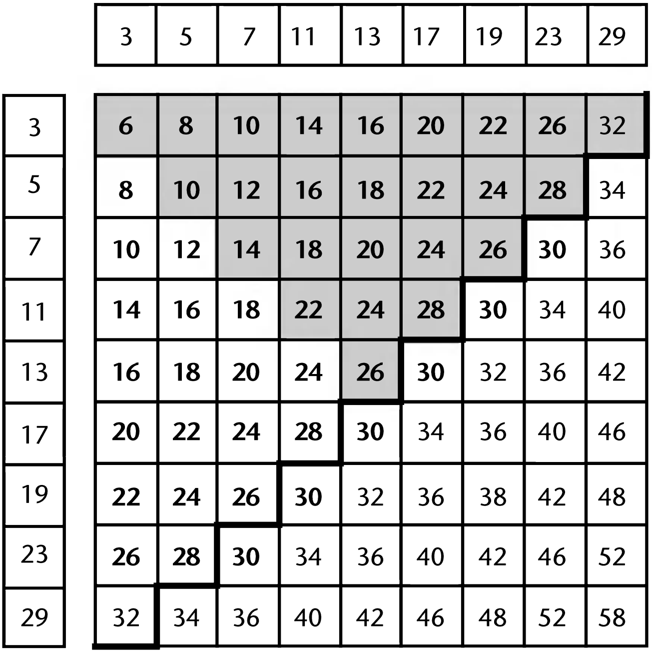
---

[1] Доксиадис А. Дядя Петрос и проблема Гольдбаха. — М.: АСТ, 2002.

содержит 1 + 4, 2 + 4, 3 + 4, 1 + 5, 2 + 5, 3 + 5, т. е. {5, 6, 7, 8}. Некоторые числа возникают здесь не по одному разу; к примеру, 6 = 2 + 4 = 1 + 5. Я называю подобные повторы перекрытием.

Теперь можно сформулировать бинарную гипотезу Гольдбаха заново: если сложить множество простых чисел с самим собой, то полученное в результате множество будет содержать все четные числа больше двух. Такое изменение формулировки может показаться немного банальным — так оно, кстати, и есть, — но оно помогает переместить проблему в ту область математики, где есть некоторые убедительные теоремы общего характера. Немного мешает число 2, но от него можно без труда избавиться. 2 — единственное целое простое число, и при сложении его с любым другим простым числом результат получается нечетный. Так что во всем, что касается гипотезы Гольдбаха, о двойке можно просто забыть. Однако 2 + 2 нам потребуется для представления числа 4, поэтому нам придется ограничить свое внимание четными числами начиная с 6.

В качестве эксперимента рассмотрим простые числа до 30 включительно. Таких чисел девять: {3, 5, 7, 11, 13, 17, 19, 23, 29}. При сложении этого множества с самим собой получится то, что можно увидеть на рис. 3: я выделил суммы, меньшие или равные 30 (диапазон четных чисел, в который укладываются все простые до 29) жирным шрифтом. При таком представлении результата ясно видны две простые закономерности. Во-первых, вся таблица симметрична относительно главной диагонали, поскольку $a + b = b + a$. И, во-вторых, выделенные числа занимают приблизительно левую верхнюю половину таблицы (см. рис. 3) над жирной (проходящей по диагонали) линией. Мало того, в середине они даже норовят вылезти за нее. Происходит это потому, что в среднем большие простые числа встречаются реже, чем маленькие. Дополнительная выпуклость посередине с лихвой компенсирует числа 32 в верхнем правом и нижнем левом углах.



	3	5	7	11	13	17	19	23	29
3	**6**	**8**	**10**	**14**	**16**	**20**	**22**	**26**	32
5	**8**	**10**	**12**	**16**	**18**	**22**	**24**	**28**	34
7	**10**	**12**	**14**	**18**	**20**	**24**	**26**	**30**	36
11	**14**	**16**	**18**	**22**	**24**	**28**	**30**	34	40
13	**16**	**18**	**20**	**24**	**26**	**30**	32	36	42
17	**20**	**22**	**24**	**28**	**30**	34	36	40	46
19	**22**	**24**	**26**	**30**	32	36	38	42	48
23	**26**	**28**	**30**	34	36	40	42	46	52
29	32	34	36	40	42	46	48	52	58

Рис. 3. Суммы пар простых чисел до 30. Жирным шрифтом выделены суммы, меньшие или равные 30. Жирной линией отмечена диагональ. Серый фон: исключение симметричных пар. Затененная область занимает чуть больше четверти квадрата

Теперь мы можем сделать некоторые грубые оценки. Я мог бы быть более точным, но этого вполне достаточно. Число ячеек в таблице составляет 9 × 9 = 81. Около половины чисел в этих ячейках находятся в левом верхнем треугольнике. Благодаря симметрии все числа, кроме лежащих на диагонали, имеют симметричную пару, так что число независимых ячеек составляет примерно 81/4, т. е., округляя, 20. В интервале от 6 до 30 содержится 13 четных чисел, поэтому 20 (и даже больше) выделенных чисел могут принимать лишь 13 четных значений. Это значит, что в данном диапазоне потенциальных сумм двух простых больше, чем четных чисел. Представьте, что вы на ярмарке и вам нужно 20 мячиками поразить 13 мишеней. Согласитесь, что шанс попасть в большую часть из них у вас будет неплохой. Тем не менее по нескольким вы можете и промазать. Иными словами, не исключено, что некоторых четных чисел все же будет не хватать.

В данном случае все числа на месте, но практические аргументы такого рода не позволяют полностью исключить подобную возможность. Однако из этого примера видно, что перекрытий должно быть немало: ведь одни и те же выделенные числа встречаются в интересующей нас четверти таблицы по несколько раз. Почему? Потому что 20 сумм должны уложиться в множество, где всего 13 членов. Поэтому каждое выделенное число в среднем встречается в таблице 1,5 раза. (Реальное количество сумм — 27, и более точная оценка показывает, что каждое выделенное число встречается дважды.) Если же каких-то четных чисел в таблице не хватает, то перекрытие должно быть еще больше.

Можно сыграть в ту же игру в более широком диапазоне, с более высоким верхним пределом — скажем, до одного миллиона. Формула, известная как теорема о распределении простых чисел (см. главу 9), дает нам возможность подсчитать количество простых чисел в интервале до любого заданного числа x. Эта оценка — $x/\log x$. В интервале до 1 000 000 количество простых оценивается по этой формуле в 72 380. (Точное их число 78 497.) Серый фон занимает около четверти соответствующей таблицы, поэтому в нем примерно $n^2/4 = 250$ млрд выделенных чисел — столько в этом диапазоне возможных сумм двух простых. Это намного больше, чем количество четных чисел в этом же диапазоне (их полмиллиона). Теперь перекрытие должно быть гигантским, а суммы должны возникать в среднем по 500 000 раз каждая. Так что шанс на то, что какое-то четное число окажется пропущено, многократно снижается.

Приложив еще некоторые усилия, мы можем с помощью этого метода оценить вероятность того, что некое четное число в заданном диапазоне не окажется суммой двух простых, исходя из того, что простые числа распределяются случайно с периодичностью, описываемой теоремой о распределении простых чисел, т. е. что в диапазоне до любого заданного x находится около $x/\log x$ простых чисел. Именно это сделали Харди и Литлвуд. Они понимали, что такой подход не является строгим, поскольку простые числа определяются достаточно специфически и распределены на самом деле не случайно. Тем не менее разумно ожидать, что реальные результаты не войдут в противоречие с этой вероятностной моделью, поскольку определяющее свойство простых чисел, судя по всему, очень слабо связано с тем, что происходит при сложении двух таких чисел.

Несколько стандартных методов в этой области математики используют примерно такой же подход, но стараются дополнительными средствами сделать свою аргументацию как можно более строгой. В качестве примера можно привести различные варианты решета, построенные на базе решета Эратосфена. Общие теоремы о плотности чисел в сумме двух множеств и возникающие в ней при очень больших множествах пропорции также оказываются весьма полезными инструментами.

В случаях, когда математическая гипотеза в конце концов находит подтверждение, ее история часто развивается по стандартному шаблону. На протяжении некоторого времени разные люди доказывают верность этой гипотезы при каких-либо ограничениях. Каждый такой результат улучшает предыдущий и снимает часть ограничений, но со временем этот путь исчерпывает свои возможности. Наконец появляется новая остроумная идея — и завершает доказательство.

К примеру, гипотеза в теории чисел может утверждать, что каждое положительное целое число может быть представлено каким-то определенным образом с использованием, скажем, шести специфических чисел (простых, квадратов, кубов, каких угодно еще). Здесь ключевыми моментами являются *каждое* положительное целое и *шесть* специфических чисел.

Первые попытки подступиться к этой проблеме дают слабые результаты, но постепенно, посредством небольших шажков, они улучшаются.

Первым шагом часто является доказательство какого-нибудь утверждения вроде, например, такого: каждое положительное целое число, которое не делится на 3 и 11, за исключением некоторого конечного их количества, может быть представлено через некое гигантское количество — скажем, 10^{666} — чисел оговоренного вида. Как правило, такая теорема умалчивает о том, сколько и каких существует исключений, так что результат невозможно приложить непосредственно к любому заданному целому числу. Следующий шаг состоит в том, чтобы обозначить границы эффективности, т. е. доказать, что каждое целое число больше 10^{1042} может быть представлено таким образом. Затем снимается ограничение по делимости на 3, а немного позже и на 11. После этого авторы один за другим начинают снимать ограничения: одни уменьшают число 10^{666}, другие 10^{1042}, третьи — то и другое одновременно. Типичным улучшением может быть, к примеру, такое: каждое целое число больше $5{,}8 \times 10^{17}$ может быть представлено с использованием не более 4298 чисел оговоренного вида.

Тем временем другие исследователи продвигаются снизу вверх, начиная с маленьких чисел, и доказывают, часто при помощи компьютерных расчетов, что, скажем, каждое число, меньшее или равное 10^{12}, может быть выражено с использованием не более шести тех самых чисел. Примерно за год 10^{12} превращается (за пять последовательных шагов, усилиями разных исследователей или групп) в $11{,}0337 \times 10^{29}$. Следует отметить, что ни один из перечисленных шагов не является ни рутинным, ни простым; напротив, они совершаются с привлечением хитроумных специальных методов, которые ничего не говорят о более общем подходе, и доказательство при каждом последовательном шаге становится все более сложным и длинным. Через несколько лет такого постепенного продвижения это число при помощи примерно тех же идей, но более мощных компьютеров и новых ухищрений удается поднять до 10^{43}. На этом, однако, метод стопорится, и все сходятся во мнении, что никакие уловки не помогут таким способом доказать полный вариант.

Гипотеза пропадает из виду, над ней уже никто не работает. Бывает, что продвижение почти совсем останавливается. Иногда без новостей проходит лет 20… И вдруг, как гром среди ясного неба, какие-нибудь Чизбургер и Чипс заявляют, что им удалось получить полное доказательство, переформулировав гипотезу в терминах комплексных метаэргодических квазимножеств и приложив теорию византийского квислинга. После нескольких лет споров о тонких моментах логики и затыкания нескольких дыр в доказательстве математическое сообщество признает его корректным и немедленно задается вопросами, не существует ли более простого способа получить тот же результат и нельзя ли его улучшить.

В последующих главах вы не раз увидите эту схему в действии. Но если рассказывать обо всем этом подробно, то может получиться довольно скучно, поэтому я не буду перечислять всех, кому удалось более точно определить экспоненту в гипотезе Джекила–Хайда, выяснив, что это не 1,773, а 1,771 + e для любого положительного e (как бы ни гордились Баггинс и Крумм своим последним достижением на этой ниве). Я опишу несколько значимых вкладов, оставив все другие за скобками. И дело не в том, что работа Баггинса и Крумма кажется мне незначительной. Может быть, она даже вымостила дорогу к прорывному открытию Чизбургера–Чипса. Но, по правде говоря, только специалисты, внимательно следящие за развитием событий, могут затаив дыхание ждать следующего крошечного шажка.

Поэтому в будущем я буду опускать некоторые подробности, но сейчас давайте посмотрим, как развивался процесс в случае с проблемой Гольдбаха.

Уже доказаны некоторые теоремы, помогающие продвинуться по пути решения проблемы Гольдбаха. Первый серьезный прорыв произошел в 1923 г., когда Харди и Литлвуд при помощи своих аналитических методов доказали тернарную гипотезу Гольдбаха для всех достаточно больших нечетных чисел. Однако их доказательство опиралось на другую великую проблему — обобщенную гипотезу Римана, о которой мы поговорим в главе 9. Эта проблема до сих пор остается нерешенной, так что в доказательстве Харди и Литлвуда есть существенный пробел. В 1930 г. Лев Шнирельман сумел заполнить этот пробел при помощи замысловатого варианта их собственных рассуждений, основанных на методах решета. Он доказал, что ненулевая доля всех чисел может быть представлена в виде суммы двух простых. Добавив к этому результату некоторые общие рассуждения о сложении последовательностей, он доказал, что существует такое целое число C, что любое натуральное число есть сумма не более C простых чисел. Это число получило известность как постоянная Шнирельмана. В 1937 г. аналогичные результаты получил Иван Виноградов, но его метод также не позволял сказать конкретно, насколько велики «достаточно большие» числа. В 1939 г. Константин Бороздин доказал, что они начинаются не позже чем с числа $3^{14\,348\,907}$. К 2002 г. Лю Минчит и Ван Тяньцзэ снизили границу «достаточно больших чисел» до e^{3100}, что равняется примерно 2×10^{1346}. Это число гораздо меньше, но все же слишком велико для того, чтобы все нижележащие числа можно было проверить перебором на компьютере.

В 1969 г. Николай Климов сумел установить, что постоянная Шнирельмана не превышает 6 млрд. Другим математикам удалось сделать более точную оценку, и в 1982 г. Ханс Ризель и Роберт Воган снизили эту цифру до 19. Хотя 19, разумеется, многим лучше 6 млрд, все признаки указывают на то, что на самом деле постоянная Шнирельмана равняется всего лишь 3. В 1995 г. Лешек Каницкий снизил верхний предел до 6 в общем случае и до 5 для нечетных чисел, но ему тоже пришлось предположить истинность гипотезы Римана. Его результаты вместе с численной проверкой гипотезы Римана вплоть до 4 × 10^{14}, которую осуществил Йорг Рихштейн, доказали бы, что постоянная Шнирельмана не превосходит 4, но опять же при условии истинности гипотезы Римана. В 1997 г. Жан-Марк Дезуйе, Гоув Эффингер, Херман те Риле и Дмитрий Зиновьев показали, что из обобщенной гипотезы Римана (см. главу 9) следует тернарная гипотеза Гольдбаха. Иными словами, каждое нечетное число, за исключением 1, 3 и 5, является суммой трех простых чисел.

Поскольку на данный момент гипотеза Римана не доказана, имеет смысл постараться снять это условие. В 1995 г. французский математик Оливье Рамаре снизил верхнюю оценку для представления нечетных чисел до 7 без использования гипотезы Римана. Более того, он доказал более сильное утверждение: каждое четное число является суммой не более чем шести простых чисел. (Чтобы разобраться с нечетными числами, вычтем из любого нечетного 3: результат четный, поэтому он является суммой шести или менее простых. Первоначально взятое нечетное есть эта сумма плюс простое число 3, т. е. для его получения требуется не более семи простых.) Главным прорывом стало уточнение существующих оценок для некоторой части чисел определенного диапазона до двух: эти числа являются суммой двух простых. Ключевой результат Рамаре состоит в том, что для любого числа n больше e^{67} (это примерно 1,25 × 10^{29}) по крайней мере пятая часть чисел, лежащих между n и $2n$, является суммой двух простых. Далее при помощи методов решета и теоремы Ганса-Генриха Остманна о суммах последовательностей, доработанной Дезуйе, можно доказать, что каждое четное число, большее 10^{30}, есть сумма максимум шести простых чисел.

Остается разобраться лишь с промежутком между 4 × × 10^{14}, до которого Йорг Рихштейн проверил теорему численно при помощи компьютера, и 10^{30}. Как часто бывает, эти числа слишком велики для непосредственной компьютерной проверки, поэтому Рамаре доказал целую серию специализированных теорем о количестве простых чисел в небольших интервалах.

Эти теоремы опираются на истинность гипотезы Римана в определенных пределах, что можно проверить при помощи компьютера. Так что доказательство состоит преимущественно из концептуальных теоретических рассуждений с привлечением компьютера для решения этой узкой задачи. Рамаре закончил свою статью указанием на то, что при помощи аналогичного подхода в принципе можно было бы снизить число простых с 7 до 5. Однако на этом пути возникают очень серьезные практические препятствия, и он написал, что такое доказательство «невозможно провести при помощи современных компьютеров».

В 2012 г. Теренс Тао преодолел эти препятствия, используя в корне другой подход. Он разместил в Интернете статью, которая в настоящий момент (когда я пишу все это) рассматривается для публикации. Основу работы составляет следующая теорема: каждое нечетное число можно представить в виде суммы не более чем 5 простых чисел. Это снижает постоянную Шнирельмана до 6. Тао получил известность благодаря своей способности решать сложные проблемы в самых разных областях математики. Его доказательство использует для решения проблемы несколько мощных методик и требует привлечения компьютеров. Если число 5 в теореме Тао удалось бы снизить до 3, то тернарная гипотеза Гольдбаха была бы доказана, а верхняя граница для постоянной Шнирельмана снижена до 4. Тао подозревает, что сделать это возможно, но нужны новые идеи.

Бинарная гипотеза Гольдбаха представляется еще сложнее. В 1998 г. Дезуйе, Саутер и те Риле проверили ее для всех четных чисел вплоть до 10^{14}. К 2007 г. Томаш Оливейра-и-Сильва улучшил этот результат до 10^{18} и продолжает расчеты. Мы знаем, что каждое четное целое число можно представить в виде суммы не более чем шести простых чисел — это доказал Рамаре в 1995 г. В 1973 г. Чэнь Цзинжунь доказал, что каждое достаточно большое четное целое может быть представлено в виде суммы простого и полупростого (это либо простое число, либо произведение двух простых) чисел. Близко, но не то. Тао заявил, что бинарную гипотезу Гольдбаха невозможно доказать при помощи его методов. Сложение трех простых чисел дает гораздо большее перекрытие результатов в том смысле, в каком мы говорили о перекрытии при обсуждении рис. 3, чем сложение двух простых, фигурирующих в бинарной гипотезе Гольдбаха, а методы и Тао, и Рамаре неоднократно используют это свойство.

Итак, через несколько лет мы, возможно, получим полное доказательство тернарной гипотезы Гольдбаха, из которой, в частности, следует, что каждое четное число можно представить в виде суммы не более чем четырех простых[1]. Но бинарная гипотеза Гольдбаха, вероятно, будет по-прежнему ставить математиков в тупик.

За 2300 лет, прошедших с момента, когда Евклид доказал несколько базовых теорем о простых числах, мы узнали о них немало. Однако остается еще очень много того, чего мы по-прежнему не знаем.

К примеру, мы знаем, что существует бесконечно много простых чисел вида $4k + 1$ и $4k + 3$. В более общем виде это утверждение выглядит так: любая арифметическая прогрессия $ak + b$ с постоянными параметрами a и b содержит бесконечно много простых чисел, если a и b не имеют общих делителей. К примеру, пусть $a = 18$. Тогда $b = 1, 5, 7, 11, 13$ или 17. Следовательно, существует бесконечно много простых чисел видов $18k + 1$, $18k + 5$, $18k + 7$, $18k + 11$, $18k + 13$ или $18k + 17$. Но это неверно для $18k + 6$, например, потому что 18 кратно 6. Ни одна арифметическая прогрессия не может состоять *только* из простых чисел, но недавний серьезный прорыв — теорема Грина–Тао — показывает, что последовательность простых чисел содержит арифметические прогрессии произвольной длины. В 2004 г. Бен Грин и Теренс Тао разработали очень глубокое и сложное доказательство этого утверждения, что внушает надежду: на самые слож-

---

[1] Тернарная гипотеза Гольдбаха была окончательно доказана перуанским математиком Харальдом Гельфготтом в 2013 г. — *Прим. ред.*

ные вопросы, какими бы неприступными они ни выглядели, в конце концов может быть получен ответ.

Снимаем шляпу, а потом надеваем ее — и вновь за работу: мы немедленно задаемся вопросом о более сложных формулах с k. Не существует простых чисел вида k^2; не существует и простых вида $k^2 - 1$, за исключением 3, поскольку подобные выражения раскладываются на множители. Однако выражение $k^2 + 1$ не имеет очевидных делителей, и простых чисел такого вида можно найти множество:

$$2 = 1^2 + 1,\ 5 = 2^2 + 1,\ 17 = 4^2 + 1,\ 37 = 6^2 + 1 \text{ и т. д.}$$

Можно привести пример и с бо́льшими цифрами, хотя особого смысла в этом нет:

$$18\,672\,907\,718\,657 = (4\,321\,216)^2 + 1.$$

Предполагается, что таких простых чисел тоже бесконечно много, но до сих пор не доказано ни одного подобного утверждения ни для одного конкретного многочлена, в котором k стояло бы в степени выше единицы. Очень правдоподобное предположение сделал в 1857 г. Виктор Буняковский: любой многочлен от k, не имеющий очевидных делителей, представляет бесконечное множество простых чисел. Исключение составляют не только разложимые многочлены, но и такие многочлены, как $k^2 + k + 2$ (этот многочлен всегда делится на 2, хотя и не имеет алгебраических делителей).

Некоторые многочлены, судя по всему, обладают особыми свойствами. Классический пример: $k^2 + k + 41$. Этот простое число, если $k = 0, 1, 2, \ldots, 40$ и, строго говоря, если $k = -1, -2, \ldots, -40$ тоже. Длинные цепочки простых чисел при последовательных значениях k попадаются редко, и о них мы кое-что знаем. Но в целом вся эта область весьма загадочна.

Гипотеза о парах простых чисел почти так же знаменита, как гипотеза Гольдбаха, и, судя по всему, столь же неприступна.

Вот ее суть: существует бесконечно много пар простых чисел с разницей в 2. Приведем несколько примеров:

3 и 5, 5 и 7, 11 и 13, 17 и 19.

На сегодняшний день (на январь 2012 г.) наибольшими известными парными простыми являются числа $3\,756\,801\,695\,685 \times 2^{666\,669} \pm 1$, содержащие по 200 700 десятичных знаков. Они были найдены в 2011 г. в рамках проекта распределенных вычислений PrimeGrid. В 1915 г. Вигго Брун при помощи одного из вариантов решета Эратосфена доказал, что сумма чисел, обратных всем парным простым, сходится, в отличие от суммы чисел, обратных всем простым. В этом смысле парные простые встречаются относительно редко. При помощи аналогичных методов он доказал также, что существует бесконечно много целых n, таких, что n и $n + 2$ имеют не больше девяти простых делителей. Харди и Литлвуд при помощи своих эвристических методов пришли к выводу, что количество пар простых, меньших x, асимптотически приближается к

$$2a\frac{n}{(\log n)^2},$$

где a — константа, равная приблизительно 0,660161. Идея в том, что в данном случае можно считать простые числа возникающими случайно с частотой, которая делает общее число простых вплоть до x приблизительно равным $x/\log x$. Аналогичных гипотез и эвристических формул существует множество, но строгих доказательств для них опять же не существует.

Да, в математике есть сотни открытых вопросов, имеющих отношение к простым числам. Одни из них просто любопытны, другие глубоки и имеют большую важность. С некоторыми вопросами из последней категории нам еще предстоит встретиться в главе 9. Ведь несмотря на все успехи математики за последние 2500 лет, скромные простые числа не потеряли ни своей притягательности, ни загадочности.

3

Тайна числа π

Квадратура круга

Простые числа известны давно, но круг — еще более древнее понятие. И именно он породил великую задачу, на решение которой ушло больше 2000 лет. Речь идет об одной из взаимосвязанных геометрических задач, корни которых уходят глубоко в античные времена. Главное действующее лицо этой истории — число π (греческая буква «пи»), знакомое нам по школьной программе в связи с окружностями и сферами. Численно это число равно 3,14159 и еще чуть-чуть; нередко также используется приблизительное значение 22/7. Десятичные знаки в записи π никогда не заканчиваются и не повторяются в одной и той же последовательности снова и снова. Нынешний рекорд вычисления точного значения числа π составляет 10 трлн знаков после запятой. Этот результат Александр Йи и Шигеру Кондо опубликовали в октябре 2011 г. Расчеты такого рода важны как способ испытания быстрых компьютеров или новых, еще более хитроумных методик вычисления числа π, но от численного результата как такового почти ничего не зависит. Причина интереса к числу π не в том, что без него невозможно вычислить длину окружности. Это странное число то и дело мелькает в самых разных областях математики, причем не только в формулах, имеющих отношение к кругам и сферам, и заводит в невероятные дебри. Тем не менее школьные формулы тоже важны, к тому же они отражают древнегреческое происхождение π.

Там одной из величайших проблем считалась нерешенная задача о квадратуре круга. В современном языке эта фраза часто используется иносказательно и означает безнадежное, бессмысленное или тщетное предприятие. Как многие общеупотребительные фразы, берущие начало в научной терминологии, эта с течением времени не раз меняла значение[5]. В греческие времена попытка найти квадратуру круга представлялась вполне разумным начинанием. Разница формы этих фигур — прямые или изогнутые границы — никакого значения не имеет: многие аналогичные задачи решаются[6]. Однако со временем выяснилось, что эта конкретная задача не может быть решена заданными методами. Чтобы это доказать, пришлось проявить изобретательность и сделать серьезные теоретические выкладки, но общую идею доказательства понять все же можно.

В математике под квадратурой круга понимают построение квадрата, равного *по площади* данному кругу, при помощи традиционных евклидовых методов. Вообще говоря, греческая геометрия допускала и другие методы, поэтому важно сразу определить, какие из них следует использовать. Но тогда неразрешимость задачи говорит только об ограниченности выбранных методов; из нее не следует, что мы не в состоянии определить площадь круга. Просто делать это придется иначе. Доказательство неразрешимости задачи о квадратуре круга помогает понять, почему греческие геометры не смогли найти требуемое построение: его просто не существует. Если разобраться, то именно поэтому им пришлось прибегать к довольно странным, чуть ли не эзотерическим методам. Так что окончательное решение этой задачи, хотя и отрицательное, помогло ученым прояснить довольно серьезную историческую загадку. Оно позволило также больше не терять времени на поиски несуществующего построения — хотя, к сожалению, всегда найдутся те, кто не сможет или не захочет принять окончательный результат, как бы тщательно его ни разжевывали[7].

В «Началах» Евклида для геометрических построений используются идеальные версии двух математических инструментов: линейки и циркуля. (Поскольку у циркуля две ножки, про него, вероятно, следовало бы говорить *циркули,* — ведь бумагу мы режем ножницами, а не одним ножницем, но я буду пользоваться традиционной терминологией.) При помощи этих инструментов геометры «чертят» на умозрительном листе бумаги — евклидовой плоскости.

Форма инструментов определяет их возможности. Циркуль представляет собой два прямых жестких стержня, соединенных шарниром. Конец одного стержня заострен, на конце другого закреплен заостренный грифель. При помощи циркуля можно нарисовать круг или часть круга определенного радиуса с центром в определенной точке. Линейка еще проще: у нее есть прямой край, по которому можно провести прямую линию. В отличие от линеек, которые сегодня можно купить в любом канцелярском магазине, линейка Евклида не имеет разметки, и это важное ограничение для математического анализа ее возможностей.

Почему речь идет об идеальных версиях инструментов, понятно: считается, что с их помощью проводятся бесконечно тонкие линии. Более того, все прямые получаются идеально прямыми, а окружности — идеально круглыми. Бумага также идеально плоская и ровная. Еще один ключевой элемент евклидовой геометрии — представление об идеальной точке. Точка ставится на бумаге, но физически такая точка невозможна: она не имеет размера. «Точка, — говорит Евклид в первой фразе своих “Начал”, — это то, что не имеет частей». По описанию она немного напоминает атом или, если вы немного в курсе современной физики, элементарную частицу, но в сравнении с геометрической точкой и атом, и частица — гигантские объекты. Однако в рамках обыденных представлений идеальная точка Евклида, атом и карандашная точка на бумаге одинаково хорошо годятся для геометрических построений.

В реальном мире идеал недостижим, как бы мы ни старались заточить карандаш и какой бы гладкой ни делали бумагу. Но в данном случае идеализм — достоинство, поскольку идеализация значительно упрощает математику. К примеру, пересечение двух реальных карандашных линий представляет собой небольшую размытую область в виде параллелограмма, но математические линии пересекаются исключительно в точке. Откровения, полученные из идеальных окружностей и линий, нередко можно перенести в реальный мир и применить к реальным несовершенным фигурам. Именно так работает волшебство математики.

Две точки определяют единственную прямую, которая через них проходит. Чтобы построить эту прямую, прикладываем нашу идеальную линейку так, чтобы ее сторона проходила через обе точки, и проводим вдоль нее идеальным карандашом. Две точки также определяют круг: выберите одну из точек — она станет центром окружности — и поставьте в нее острие циркуля; затем разведите ножки циркуля так, чтобы кончик грифеля встал на вторую точку. А теперь ведите грифель по дуге, аккуратно удерживая острие в центре. Две прямые определяют единственную точку пересечения — если, конечно, они не параллельны; в этом случае прямые не пересекаются, зато широко распахивается логический ящик Пандоры. Прямая и окружность определяют две точки, если пересекаются, и одну, если прямая лишь касается окружности; если окружность слишком мала, чтобы дотянуться до прямой, пересечения не будет. Точно так же две окружности могут пересекаться в двух точках, в одной или не пересекаться вовсе.

Расстояние — фундаментальная концепция, без которой немыслимо современное прочтение евклидовой геометрии. Расстояние между двумя точками измеряется по прямой, их соединяющей. Евклид, разрабатывая свою геометрию, обходился без явно выраженной концепции расстояния, но он и без этого мог определить, когда два отрезка прямой имеют *одинаковую* длину. Это очень просто: достаточно поставить ножки циркуля на концы одного отрезка, перенести инструмент ко второму отрезку и посмотреть, встанут ли ножки на его концы. Если встанут, то длины этих отрезков одинаковы; если нет — нет. Эта процедура вовсе не требует измерения реальных длин.

Из этих базовых составляющих геометры могут построить более интересные формы и конфигурации. Три точки определяют треугольник, если только не лежат на одной прямой. Две прямые, пересекаясь, образуют угол. Особенно важен прямой угол, а развернутый угол соответствует двум составленным вместе прямым углам. И так далее и тому подобное, до бесконечности. «Начала» Евклида включают в себя 13 книг и с каждой книгой все глубже зарываются в следствия этих простых начал.

Основное содержание «Начал» — теоремы, строительный материал геометрии. Кроме того, Евклид объясняет, как решать геометрические задачи при помощи «построений», сделанных с применением линейки и циркуля. Как, имея две точки, соединенные отрезком прямой, получить среднюю точку отрезка? Как разделить отрезок на три равные части? Как, имея угол, построить другой угол, равный в точности половине первого? Но некоторые простые построения неожиданно оказались неуловимыми. К примеру, трисекция угла: постройте угол, который ровно втрое меньше заданного. С отрезками такое проходит, но для углов никому так и не удалось отыскать соответствующий метод. С любой степенью приблизительности — да, пожалуйста. Построить точно при помощи циркуля и линейки — нет, увольте. Однако в реальной жизни никому обычно не надо делить угол ровно натрое, так что этот конкретный вопрос не вызвал особых проблем.

Куда больше шума наделало построение, обойтись без которого было никак нельзя: имея заданный круг, построить квадрат той же площади. Это и есть задача о квадратуре круга. С точки зрения греков, если невозможно решить эту задачу, то нельзя и утверждать, что круг вообще имеет площадь.

Ну и что, что он очевидно заключает в себе определенное пространство — площадь-то интуитивно определяется по тому, *сколько* пространства заключает в себе фигура. Евклид и его последователи, в частности Архимед, сошлись на прагматическом решении: они считали, что круг имеет площадь, но построить квадрат той же площади невозможно. О площади круга, конечно, тоже можно кое-что сказать. К примеру, можно доказать со всей логической строгостью, что площадь круга пропорциональна квадрату его диаметра. А вот что невозможно сделать, не решив задачу квадратуры круга, так это начертить отрезок, длина которого будет представлять собой коэффициент этой пропорциональности.

Греки не смогли решить задачу квадратуры круга при помощи линейки и циркуля, им пришлось удовлетвориться другими методами. Один из них воспользовался для этого кривой, получившей название квадратрисы. Судя по всему, позднейшие комментаторы сильно преувеличили значение, которое греческие геометры придавали тому, что всякое построение должно делаться только при помощи линейки и циркуля. По сути, мы даже не можем сказать наверняка, действительно ли греки считали квадратуру круга такой важной задачей. К XIX в., однако, эта проблема приобрела поистине вселенские масштабы. Математика, не способная ответить на такой простой и понятный вопрос, — все равно что повар, не способный сварить яйцо вкрутую.

Формулировка задачи — квадратура круга — звучит очень по-геометрически. Так и есть, это действительно геометрическая задача. А вот решение ее, как оказалось, лежит в области вовсе не геометрии, а алгебры. Дело в том, что решение великих задач часто основывается на выявлении неожиданных связей между разными, на первый взгляд, разделами математики. Связь геометрии и алгебры сама по себе не является чем-то беспрецедентным, но тот факт, что она имеет какое-то отношение к квадратуре круга, был замечен далеко не сразу. А потом, когда связь уже была установлена, возникли чисто технические сложности, и для их разрешения потребовалось привлечь еще один раздел математики — математический анализ. По иронии судьбы первый шаг к прорыву был сделан в четвертой области математики — в теории чисел. В результате была решена геометрическая задача, в решаемость которой греки не поверили бы даже в самых смелых своих мечтах и о которой, насколько нам известно, никогда не думали: задача о построении при помощи циркуля и линейки правильного многоугольника с 17 сторонами.

Звучит дико, особенно если добавить, что для правильных многоугольников с 7, 9, 11, 13 или 14 сторонами ничего подобного не существует, зато многоугольник с 3, 4, 5, 6, 8, 10 и 12 сторонами построить можно. Однако в данном случае за безумием скрывается система, причем такая, что ее выявление заметно обогатило математику.

Начнем с начала: что такое правильный многоугольник? Многоугольник вообще — это фигура, ограниченная прямыми линиями. Многоугольник называется правильным, если все отрезки прямых имеют одинаковую длину и пересекаются под одинаковыми углами. Самый известный пример — квадрат: все четыре его стороны имеют одинаковую длину, а все четыре угла являются прямыми. Существуют и другие фигуры — с четырьмя равными сторонами или с четырьмя равными углами: это, соответственно, ромб и прямоугольник. Только квадрат обладает обоими свойствами одновременно. Правильный трехсторонний многоугольник — это равносторонний треугольник; существуют также правильный пятиугольник, правильный шестиугольник и т. д. (рис. 4). Евклид приводит методы построения при помощи циркуля и линейки правильных многоугольников с 3, 4 и 5 сторонами. Кроме того, греки умели последовательно удваивать число сторон, выстраивая многоугольники с 6, 8, 10, 12, 16, 20 и более сторонами. Объединив методы построения правильных многоугольников с 3 и 5 сторонами, они получили правильный 15-угольник. Но на этом продвижение остановилось, и далее, на протяжении 2000 лет, на этом направлении ничего не менялось. Никто не думал, что в этом списке могут появиться многоугольники с еще каким-то числом сторон. Никто даже не задавался этим вопросом: всем казалось, что ничего больше сделать не удастся.

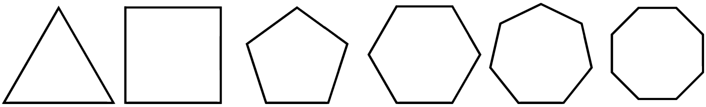


Рис. 4. Первые несколько правильных многоугольников. Слева направо: равносторонний треугольник, квадрат, правильные пятиугольник, шестиугольник, семиугольник и восьмиугольник

Понадобилось вмешательство одного из величайших математиков всех времен, чтобы обдумать немыслимое, задаться вопросами, задавать которые бесполезно, и получить поистине поразительный ответ. Имя этого математика — Карл Гаусс.

Родился Гаусс в бедной семье в городе Брауншвейге в Германии. Его мать Доротея была неграмотной и не смогла даже записать дату рождения ребенка. Однако она помнила, что было это в 1777 г., за восемь дней до праздника Вознесения. Позже Гаусс сам вычислил точную дату своего рождения при помощи разработанной им формулы расчета дат Пасхи. Отец ученого Гебхард происходил из крестьянской семьи, но зарабатывал на жизнь разной работой: копал канавы, был садовником, уличным мясником, счетоводом похоронной конторы. А их сын оказался вундеркиндом: рассказывали, что уже в трехлетнем возрасте он исправлял отцовские ошибки в арифметике. Его способности, распространявшиеся помимо математики и на языки, побудили герцога Брауншвейгского оплатить обучение мальчика в Брауншвейгском университете. Будучи студентом, Гаусс самостоятельно открыл для себя несколько важных математических теорем, доказанных знаменитыми учеными, такими как Эйлер. Однако его теорема о правильном 17-угольнике грянула как гром среди ясного неба.

К тому времени прошло уже 140 лет с тех пор, как была установлена тесная связь между геометрией и алгеброй. В приложении к «Рассуждению о методе…» Рене Декарт формализовал идею, давно витавшую в воздухе: представление о системе координат. По существу, он взял евклидову девственно чистую плоскость — пустой лист бумаги — и превратил его в лист, расчерченный на квадраты (инженеры и ученые называют такую бумагу миллиметровкой). Для начала нарисуйте на бумаге две прямые линии, горизонтальную и вертикальную. Эти линии называются осями координат. Теперь можно определить положение любой точки на плоскости, задавшись вопросом: как далеко лежит эта точка в направлении вдоль горизонтальной оси и как далеко — вдоль вертикальной (см. рис. 5 слева). Эти два числа — а они могут быть как положительными, так и отрицательными, — дают исчерпывающее описание точки и называются ее координатами.

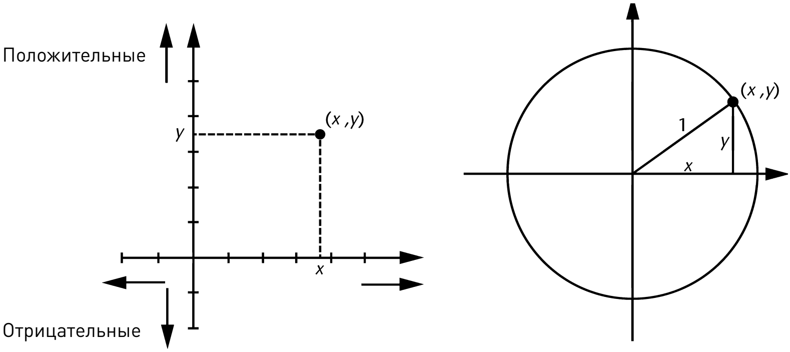


Рис. 5. Координаты на плоскости *(слева)*. Как вывести уравнение единичной окружности *(справа)*

Все геометрические свойства точек, прямых, окружностей и т. д. можно перевести в алгебраические утверждения, связанные с соответствующими координатами. Очень трудно осмысленно говорить об этих связях без использования алгебры — точно так же, как трудно говорить о футболе без использования слова «гол». Поэтому на следующих страницах мне придется привести несколько формул. Они нужны для того, чтобы показать: у главных действующих лиц этой драмы есть имена, и отношения между ними прозрачны. Согласитесь, «Ромео» — это гораздо понятнее, чем «сын итальянского патриция, полюбивший красавицу-дочь заклятого врага своего отца». Наш Ромео будет носить прозаическое имя x, а его Джульетту будут звать y.

В качестве примера того, как геометрия превращается в алгебру, рис. 5 (справа) показывает, как найти уравнение окружности единичного радиуса с центром в начале координат, где пересекаются наши две оси. Отмеченная точка имеет координаты (x, y), так что у прямоугольного треугольника на рисунке длина горизонтальной стороны равна x, а вертикальной — y. Самая длинная сторона треугольника представляет собой радиус окружности и, соответственно, равняется единице. Теорема Пифагора гласит, что сумма квадратов двух координат равняется 1. В символьном виде это звучит так: точка с координатами x и y лежит на окружности тогда и только тогда, когда ее координаты удовлетворяют условию $x^2 + y^2 = 1$. Символьная характеристика окружности получилась краткой и точной; она наглядно показывает, что речь в данном случае действительно идет об алгебре. И наоборот, любая алгебраическая характеристика пары чисел, любое уравнение с участием x и y можно интерпретировать как геометрическое утверждение о точках, прямых, окружностях или более сложных кривых[8].

Фундаментальные алгебраические уравнения включают, в частности, многочлены — комбинации различных степеней неизвестной величины x, где каждая степень x умножается на некое число, называемое коэффициентом. Наибольшая степень x есть степень многочлена. К примеру, уравнение

$$x^4 - 3x^3 - 3x^2 + 15x - 10 = 0$$

содержит многочлен, начинающийся с x^4, т. е. четвертой степени. Коэффициенты здесь равны 1, –3, –3, 15 и –10. У этого уравнения четыре различных решения: $x = 1, 2,\ 5$ и 5. Для этих чисел левая часть уравнения равняется нулю, т. е. правой части. Многочлены первой степени, такие как $7x + 2$, называются линейными и содержат только первую степень неизвестного. Уравнения второй степени, такие как $x^2 - 3x + 2 = 0$, называются квадратными и содержат вторую степень неизвестного — квадрат. Уравнение окружности содержит вторую переменную y. Однако, если у нас есть второе уравнение, связывающее x и y, к примеру уравнение какой-нибудь прямой, мы можем выразить в нем y через x и преобразовать уравнение окружности так, чтобы оно содержало только x. Это новое уравнение говорит нам о том, где прямая пересекается с окружностью. В данном случае новое уравнение является квадратным и имеет два решения. Так алгебра отражает геометрию, в которой прямая пересекает окружность в двух вполне конкретных точках.

Это свойство алгебры очень существенно влияет на геометрические построения, сделанные при помощи линейки и циркуля. Любое подобное построение, каким бы сложным оно ни было, состоит из последовательности простых шагов. Каждый шаг дает новые точки в местах, где пересекаются две прямые, две окружности или прямая и окружность. Каждая из этих прямых и окружностей определяется ранее построенными точками. Переводя геометрию на язык алгебры, можно доказать, что алгебраическое уравнение, соответствующее пересечению двух прямых, обязательно линейное, а пересечению прямой и окружности или двух окружностей — квадратное. Причина в том, что уравнение окружности содержит x^2, но не содержит более высоких степеней x. Поэтому каждый отдельный этап построения соответствует решению уравнения первого или второго порядка, не выше.

Более сложные построения представляют собой последовательности этих базовых операций. Некоторое количество алгебраических преобразований позволяет нам сделать вывод, что каждая координата любой точки, которую можно получить геометрическим построением при помощи линейки и циркуля, является решением полиномиального уравнения с целыми коэффициентами, степень которого представляет собой одну из степеней двойки. Это значит, что степень уравнения должна быть равна одному из чисел 1, 2, 4, 8, 16 и т. д.[9] Это необходимое условие существования такого построения. При должном старании из него можно извлечь точную характеристику, которой должен обладать правильный многоугольник, чтобы его можно было построить. Внезапно из сложной геометрической паутины появляется на свет аккуратное алгебраическое условие, причем применимое *к любому построению*. Необязательно даже знать, что при этом строится: достаточно, чтобы при построении использовались только линейка и циркуль.

Гаусс был знаком с этой элегантной идеей. Он знал также (к такому выводу пришел бы любой компетентный математик), что вопрос о том, какой правильный многоугольник можно построить при помощи линейки и циркуля, сводится к частному случаю, в котором многоугольник имеет простое число сторон. Чтобы понять, почему так происходит, представьте себе составное число, к примеру 15, т. е. 3 × 5. Любое гипотетическое построение правильного 15-угольника автоматически даст нам правильный же треугольник (возьмите каждую пятую вершину) и пятиугольник (каждую третью), как на рис. 6. Приложив еще немного усилий, можно так скомбинировать построения для трех- и пятиугольников, чтобы получить в результате 15-угольник[10]. Числа 3 и 5 — простые, и к ним приложима та же общая идея. Так что Гаусс сосредоточился на многоугольниках с простым числом сторон и задался вопросом о том, на что похоже нужное уравнение. Ответ оказался удивительно изящным. Так, построение правильного пятиугольника эквивалентно решению уравнения $x^5 - 1 = 0$. Замените 5 любым другим простым числом — и соответствующее утверждение тоже будет истинным.

Степень этого многочлена — 5, и это не степень двойки, о которой я говорил; тем не менее построить правильный пятиугольник можно. Гаусс быстро определил, почему: это уравнение раскладывается на две части — первого и четвертого порядка. И 1, и 4 являются степенями двойки; оказывается к тому же, что ведущую роль здесь играет уравнение четвертой степени. Чтобы понять, почему нам следует связать это уравнение с геометрией, придется привлечь новый тип числа, который обходит вниманием школьная математика, но без которого на любом более высоком уровне не обойтись. Речь идет о комплексных числах; их определяющим свойством является то, что в системе комплексных чисел из –1 можно извлечь квадратный корень.

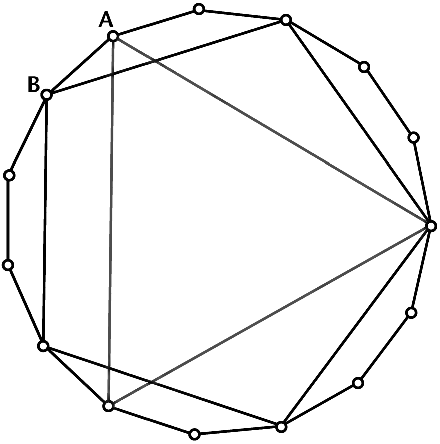


Рис. 6. Построение равностороннего треугольника и правильного пятиугольника на базе правильного 15-угольника. Чтобы восстановить алгоритм обратного построения, обратите внимание на то, что точки A и B стоят в правильном 15-угольнике рядом

Обычное «действительное» число может быть положительным и отрицательным, но его квадрат в том и другом случае положителен, так что –1 не может быть квадратом какого бы то ни было действительного числа. В некоторых случаях это сильно мешает; математики даже изобрели новый тип «воображаемого», или «мнимого», числа, квадрат которого равен –1. Нужно было как-то обозначить это новое число, для чего воспользовались первой буквой слова *imaginary* (воображаемый) — i. Обычные алгебраические операции — сложение, вычитание, умножение, деление — привели к возникновению комбинаций действительных и мнимых чисел, таких как $3 + 2i$. Такие числа называют комплексными, что вовсе не означает «сложные», а просто указывает на то, что они состоят из двух частей: 3 и $2i$. Если действительные числа располагаются на известной числовой прямой, как числа на линейке, то комплексные числа лежат на числовой плоскости, на которой мнимая ось располагается под прямым углом к действительной, а вместе они образуют систему координат (см. рис. 7, слева).

Уже 200 лет математики считают комплексные числа фундаментальной концепцией своей науки. Мы сегодня признаем, что логически комплексные числа имеют ту же основу, что и более знакомые «действительные» — ведь те тоже, подобно всем математическим структурам, представляют собой абстрактное понятие, а не реальную физическую вещь. Комплексные числа широко использовались еще до Гаусса, но их статус оставался неясным, пока Гаусс и другие математики не сорвали с них завесу тайны, раскрыв неожиданную и парадоксальную причину их привлекательности: несмотря на загадочность и неясный смысл, комплексные числа ведут себя гораздо лучше действительных. Они внесли недостающую составляющую, которой не хватало действительным числам, — обеспечили любому алгебраическому уравнению полный набор решений.

Простейший пример — квадратные уравнения. Одни из них имеют по два действительных решения, другие — не имеют ни одного. К примеру, решениями уравнения $x^2 – 1 = 0$ являются 1 и –1, а уравнение $x^2 + 1 = 0$ решений не имеет. Промежуточное положение занимает $x^2 = 0$, единственное решение которого равно 0, но в некотором смысле это единственное решение «повторяется дважды»[11]. Если же мы разрешим комплексные

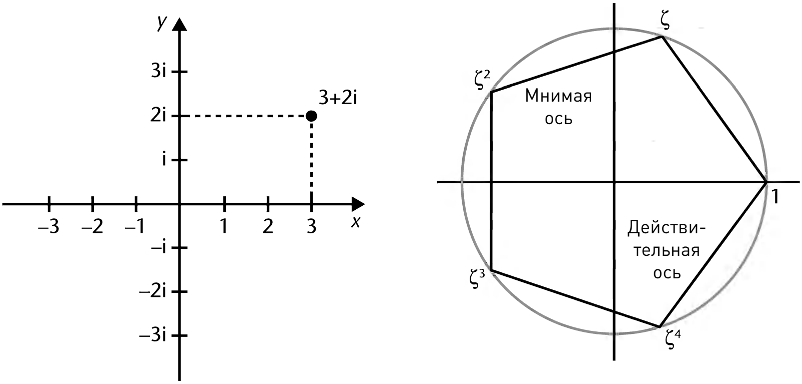


Рис. 7. Комплексная плоскость *(слева)*. Пять комплексных корней единицы *(справа)*

решения, то окажется, что $x^2 + 1 = 0$ тоже имеет два решения: i и $-i$. Гаусс не стеснялся пользоваться комплексными числами; мало того, его докторская диссертация содержала первое логически безупречное доказательство фундаментальной теоремы алгебры: число комплексных решений любого полиномиального уравнения (если корректно посчитать кратные корни) равняется степени уравнения. Поэтому квадратные уравнения (второй степени) всегда имеют по два комплексных решения, кубические (третьей степени) — по три и т. д.

Уравнение $x^5 - 1 = 0$, определяющее, как я уже сказал, правильный пятиугольник, — это уравнение пятой степени, поэтому и комплексных решений у него пять. Действительное решение одно: $x = 1$. Где же остальные четыре? Они представляют собой четыре вершины правильного пятиугольника на комплексной плоскости, притом что $x = 1$ — это пятая вершина (см. рис. 7, справа). Это соответствие — удачный пример математической красоты: элегантная геометрическая фигура становится элегантным уравнением.

Вспомним, однако, о том, что эти пять точек являются решениями уравнения пятой степени, а 5 — это не степень двойки. Но, как уже говорилось, уравнение пятой степени раскладывается на две части со степенями 1 и 4; эти части называют его неприводимыми делителями.

$$x^5 - 1 = (x - 1)\ (x^4 + x^3 + x^2 + x + 1).$$

(«Неприводимость» означает, что у этих многочленов уже нет делителей, как у простых чисел.) Первый делитель дает единственное действительное решение $x = 1$. Второй делитель дает четыре комплексных решения — и четыре вершины пятиугольника. Так что с комплексными числами все выглядит гораздо разумнее и элегантнее.

Часто сложно понять, каким путем математики прошлого приходили к новым открытиям, потому что в те времена было принято представлять только конечный результат размышлений и оставлять в стороне все ошибочные шаги, которые были сделаны в ходе исследования. Эта проблема часто осложняется и тем, что естественный ход мысли в прошлом выглядел иначе, чем сегодня. Гаусс, в частности, широко известен своей склонностью заметать следы и публиковать только конечный, тщательно отшлифованный анализ. Однако в том, что касается исследований Гаусса по построению 17-угольника, материала у нас достаточно: окончательная публикация содержит достаточно ценных указаний.

Его отправная точка новизной не отличалась. И до Гаусса кое-кто из математиков понимал, что приведенный выше анализ правильных многоугольников работает в общем случае. Построение многоугольника с числом сторон n эквивалентно решению уравнения $x^n - 1 = 0$ в комплексных числах. Более того, этот многочлен раскладывается на два многочлена вида

$$(x - 1)\ (x^{n-1} + x^{n-2} + \ldots + x^2 + x + 1).$$

Опять же первый множитель дает действительное решение $x = 1$, а остальные $n - 1$ решений получаются из второго множителя. Если n нечетное, все они комплексные; если n четное, одно из них становится вторым действительным решением $x = -1$.

Однако Гаусс заметил то, что просмотрели все остальные: иногда второй делитель можно выразить через несколько квадратных уравнений. Не представить в виде произведения более простых множителей, поскольку это невозможно, а решить с использованием уравнений, коэффициенты которых решают другие уравнения. Ключевым фактором — слабым звеном всей задачи — является одно элегантное свойство алгебраических уравнений, возникающее, когда мы решаем их подобным образом одно за другим. Такой расчет всегда эквивалентен решению единственного уравнения, но, как правило, более высокой степени. Повышение степени — цена, которую мы платим за уменьшение количества уравнений. Технически эта процедура может оказаться достаточно сложной и путаной, но одно мы можем предсказать заранее: какая получится степень. Для этого достаточно перемножить степени всех последовательных многочленов.

Если все они квадратные, то результат будет $2 \times 2 \times \ldots \times 2$, т. е. степень двойки. Поэтому, если построение существует, $n - 1$ должно быть степенью двойки. Однако этого условия не всегда достаточно. Если $n = 9$, $n - 1 = 8$, а это степень двойки. Но Гаусс выяснил, что для правильного девятиугольника построения не существует, поскольку 9 — не простое число[12]. А как насчет следующего шага, на котором мы решаем систему из четырех квадратных уравнений? Степень $n - 1$ соответствующего объединенного уравнения равна $2 \times 2 \times 2 \times 2 = 16$. Тогда $n = 17$, а это простое число.

К этому моменту Гаусс, вероятно, уже понял, что наткнулся на что-то интересное, но оставался еще один технический момент, который вполне мог все испортить. Гаусс убедился, что для существования построения правильного многоугольника с простым числом сторон это простое число должно равняться степени двойки плюс 1. Получалось, что это условие необходимо для существования построения: если оно не выполняется, такого построения не существует. Однако вполне могло оказаться, что этого условия недостаточно; в самом деле, существует множество уравнений 16-й степени, которые не сводятся к системе из четырех квадратных уравнений.

Однако был и повод для оптимизма — греческие построения. Какие простые числа там фигурировали? Только три из них: 2, 3 и 5. Все они на единицу больше какой-либо степени двойки, а именно $2^0 + 1$, $2^1 + 1$ и $2^2 + 1$. Алгебра, связанная с пятиугольником, дает дополнительную пищу для размышлений. Обдумывая все это, Гаусс доказал, что многочлен 16-й степени, соответствующий правильному 17-угольнику, действительно может быть сведен к системе квадратных уравнений. Поэтому построение правильного 17-угольника при помощи линейки и циркуля обязательно должно существовать. Аналогичным методом удалось доказать, что то же верно для любого случая, когда количество сторон является простым числом, на 1 превосходящим некоторую степень двойки. Вообще, эта идея наглядно свидетельствует, насколько хорошо Гаусс понимал математические закономерности. В основе их лежат некоторые общие теоремы теории чисел, в которые я сейчас не буду вдаваться; замечу только, что все это не случайно и у этих закономерностей существуют серьезные структурные причины. Просто надо быть Гауссом, чтобы их заметить.

Гаусс не составил полного алгоритма такого построения, но вывел формулу для решений уравнения 16-й степени. А имея формулу, можно при большом желании придумать и построение[13]. Публикуя свои идеи в «Арифметических исследованиях», он опустил несколько подробностей, но заявил, что обладает полным доказательством. Это грандиозное открытие убедило Гаусса в том, что лучше посвятить жизнь математике, а не языкам. Герцог по-прежнему не оставлял Гаусса без финансовой поддержки, но молодому человеку хотелось чего-то более стабильного. Когда астроном Джузеппе Пиацци открыл первый астероид — Цереру, — ученым удалось провести всего несколько наблюдений, прежде чем новооткрытый мир скрылся в сиянии Солнца. Астрономы тревожились, что не смогут вновь найти его. Проявив чудеса изобретательности и использовав новую методику расчета орбит, Гаусс предсказал, где новооткрытое небесное тело появится вновь, и оказался прав. В результате он получил место профессора астрономии и директора Геттингенской обсерватории и оставался на этом посту до конца своих дней.

Оказалось, что 17 — не единственное новое число такого типа. На сегодня известны еще два подобных числа: $2^8 + + 1 = 257$ и $2^{16} + 1 = 65\,537$. (Еще немного алгебры — и можно показать, что степень двойки, фигурирующая в этом выражении, сама должна быть степенью двойки; в противном случае результат не будет простым.) Однако на 16 эта закономерность прекращается, и $2^{32} + 1 = 4\,294\,967\,297$, что равно $641 \times \times 6\,700\,417$, а значит, не является простым числом. Известно, что так называемые числа Ферма $2^{2^n} + 1$ не являются простыми для $n = 5, 6, 7, \ldots$ и так до 32. Известно также, что многие более крупные числа Ферма тоже не простые. Вообще, больше простых чисел Ферма пока не найдено, но вполне возможно, что они все же существуют. Известно построение для правильного 257-угольника. Один математик посвятил много лет поиску построения для 65 537-угольника — правда, эта задача представляется несколько бессмысленной, и, кроме того, в его результатах есть ошибки[14].

Итак, основной вывод из проведенного Гауссом анализа состоит в том, что правильный многоугольник может быть построен при помощи линейки и циркуля в том и только том случае, когда число его сторон представляет собой произведение степени двойки и *различных* нечетных простых чисел Ферма. В частности, правильный девятиугольник так построить нельзя. Из этого сразу следует, что по крайней мере один угол невозможно разделить натрое построением: ведь угол равностороннего треугольника равен 60°, а одна треть такого угла — это 20°.

Но, имея такой угол, несложно построить правильный девятиугольник. Следовательно, это невозможно, и общего метода трисекции угла при помощи геометрического построения не существует.

Гаусс, записывая доказательства, опустил немало подробностей, и математики не могли просто так поверить ему на слово. В 1837 г. французский математик Пьер Ванцель опубликовал полное доказательство гауссовой характеризации пригодных для геометрического построения правильных многоугольников и сделал вывод о невозможности трисекции произвольного угла при помощи линейки и циркуля. Он доказал также невозможность построения куба объемом вдвое больше данного (т. е. доказал неразрешимость еще одной древнегреческой задачи, известной как «задача об удвоении куба»).

Причина того, что задачи трисекции угла и удвоения куба оказались неразрешимыми, заключается в том, что задействованные в них длины фигурируют в неприводимых *кубических* уравнениях — уравнениях третьей степени. Раз 3 не является степенью двойки, это все решает. Однако этот метод, на первый взгляд, не работает для квадратуры круга, причем по достаточно интересным причинам. Круг единичного радиуса имеет площадь π, а сторона квадрата той же площади равна π . Геометрические построения для квадратного корня существуют, как и построения для квадратов, так что квадратура круга, по существу, сводится к тому, чтобы взять отрезок длиной 1 и построить отрезок длиной π. Конечно, если π является решением неприводимого кубического уравнения — или любого другого неприводимого уравнения, чья степень не является степенью двойки, — то метод Ванцеля доказал бы, что квадратура круга невозможна.

Однако никто не слышал ни об одном алгебраическом уравнении, решением которого было бы в точности π, не говоря уж об уравнении степени, не являющейся степенью двойки. Приближенное значение 22/7 удовлетворяет уравнению $7x - 22 = 0$, но на самом деле эта дробь немного больше π, так что это никак нам не поможет. Если бы можно было доказать, что такого уравнения не существует, — а многие подозревали, что так оно и есть, поскольку если бы уравнение существовало, то его бы нашли, — то из этого следовала бы и невозможность решения задачи квадратуры круга. К несчастью, никто не мог доказать, что такого уравнения не существует. Алгебраический статус π пребывал в состоянии неопределенности. В конце концов этот вопрос все же удалось решить, но при помощи методов, далеко выходящих за пределы не только геометрии, но и алгебры.

Чтобы разобраться в существе дела, нам придется начать с более простой концепции. В математике существует важное различие между числами, которые можно точно выразить при помощи дроби p/q, где p и q — целые числа, и числами, которые невозможно выразить таким образом. Первые называются рациональными (поскольку представляют собой отношение, т.е. *ratio*, целых чисел), а последние — иррациональными. Так, приближенное значение $\pi(22/7)$ рационально. Существуют и более точные приближения; знаменитое $355/113$ соответствует π до шестого знака после запятой. Однако известно, что никакая дробь не может выразить π точно; число π иррационально. Это свойство, о котором математики давно догадывались, первым доказал швейцарский математик Иоганн Генрих Ламберт в 1768 г. Его доказательство основано на хитрой формуле для функции тангенса в тригонометрии, где тангенс выражается в виде цепной (непрерывной) дроби — бесконечного множества обычных дробей[15]. В 1873 г. Шарль Эрмит нашел более простое доказательство, основанное на аналитических формулах, которое доказало иррациональность не только π, но и π^2. Так что π помимо всего прочего не является корнем квадратным из какого-то рационального числа.

Ламберт выдвинул и более серьезные гипотезы. В той же статье, где доказывалась иррациональность π, он предположил, что число π трансцендентно, т.е. не является решением никакого полиномиального уравнения с целыми коэффициентами. Оно выходит за рамки алгебраического выражения. Более поздние исследования доказали правоту Ламберта. Сделано это было в два этапа. Разработанный Эрмитом новый метод доказательства иррациональности подготовил площадку, намекнув, что удачной стратегией здесь может оказаться исчисление, а точнее, его более строгая версия — анализ. Но Эрмит развил эту идею дальше и нашел чудесное доказательство трансцендентности другого знаменитого и очень любопытного числа e — основания натуральных логарифмов. Численно e приблизительно равно 2,71828, и, пожалуй, оно еще важнее, чем π. Эрмитово доказательство трансцендентности волшебно, как кролик, с помпой извлекаемый фокусником из цилиндра математического анализа. Сам кролик — это сложная формула, связанная с гипотетическим алгебраическим уравнением, корнем которого, согласно первоначальному предположению, является e. При помощи алгебраических методов Эрмит доказывает, что эта формула эквивалентна некоему ненулевому целому числу. При помощи математического анализа он доказывает, что число это должно лежать между $-1/2$ и $1/2$. Поскольку единственным целым числом в этом интервале является 0, получаем противоречие. Следовательно, предположение о том, что e является решением некоего алгебраического уравнения, неверно, а значит, e трансцендентно.

В 1882 г. Фердинанд фон Линдеман несколько усовершенствовал метод Эрмита и доказал, что если ненулевое число является решением некоего алгебраического уравнения, то e в степени этого числа не является решением никакого алгебраического уравнения. Затем он воспользовался соотношением, известным еще Эйлеру и связывающим π, e и мнимое число i, — знаменитой формулой $e^{i\pi} = -1$. Предположим, что π удовлетворяет некоему алгебраическому уравнению. То же можно сказать и про $i\pi$, а по теореме Линдемана получим, что -1 не удовлетворяет никакому алгебраическому уравнению. Это очевидно неверно: –1 является решением уравнения $x + 1 = 0$. Единственный выход из этого логического противоречия заключается в том, что π не удовлетворяет никакому алгебраическому уравнению, т. е. π трансцендентно. А это означает, что задача квадратуры круга неразрешима.

Путь от евклидовой геометрии к доказательству Линдемана получился долгим и кружным. Но математики, хоть и через две с лишним тысячи лет, все же добились своего. Однако вся эта история не просто сообщает нам о невозможности квадратуры круга. Это наглядный урок того, как вообще решаются великие математические задачи. Во-первых, математикам потребовалось точно сформулировать, что они имеют в виду, говоря о «геометрическом построении». Им пришлось определить общие черты таких построений и понять, как эти черты ограничивают возможности построений. Для поиска общих свойств потребовалось связать геометрию с другой областью математики — алгеброй. Но при решении алгебраических задач, даже не самых сложных, таких как построение правильных многоугольников, не обойтись без теории чисел. Сложный случай числа π потребовал дополнительных новшеств, и задачу пришлось перенести в еще одну область математики — математический анализ.

Ни один из перечисленных шагов не был простым или очевидным. Уже после того, как основные идеи были высказаны, потребовалось еще около 100 лет, чтобы окончательно доработать доказательство. Математики, бившиеся над этой задачей, были лучшими умами своего времени, а по крайней мере один из них входит в число величайших умов всех времен. Решение подобных задач помимо глубокого понимания математики требует настойчивости и изобретательности. Иногда на это уходят годы сосредоточенных усилий, на первый взгляд, по большей части бесплодных. Но представьте, что чувствует математик, когда его настойчивость приносит плоды, и ему наконец удается расколоть крепкий орешек, над которым человечество билось не один век. Как сказал президент Джон Кеннеди в 1962 г. в одной из речей, посвященных Лунной программе, «мы решили… это сделать… не потому, что это просто, а потому, что это сложно».

Мало что в математике имеет конец, и история числа π — не исключение. И сегодня время от времени появляются поразительные новые открытия, имеющие к нему отношение. В 1997 г. Фабрис Беллар объявил, что триллионная цифра числа π в двоичном выражении — единица. Замечательным это заявление делает не собственно факт. Поразительно то, что он не вычислял ни одной из предыдущих цифр. Он просто извлек одну конкретную цифру, что называется, из воздуха.

Такой расчет оказался возможен благодаря любопытной формуле для π, которую открыли Дэвид Бэйли, Питер Боруэйн и Саймон Плафф в 1996 г. Она может показаться несколько сложноватой, но все же посмотрим:

$$\pi=\sum_{n=0}^{\infty}\frac{1}{2^{4n}}(\frac{4}{8n+1}-\frac{2}{8n+4}-\frac{1}{8n+5}-\frac{1}{8n+6}).$$

Большой знак $\sum$ означает «просуммировать» на заданном диапазоне. Здесь n изменяется от 0 до бесконечности (∞). На самом деле Беллар пользовался формулой, которую вывел сам с использованием аналогичных методов и расчет по которой ведется чуть быстрее:

$$\pi=\frac{1}{64}\times\sum_{n=0}^{\infty}\frac{(-1)^n}{2^{10n}}(-\frac{32}{4n+1}-\frac{1}{4n+3}+\frac{256}{10n+1}-$$
$$-\frac{64}{10n+3}-\frac{4}{10n+5}-\frac{4}{10n+7}+\frac{1}{10n+9}).$$

Ключевая особенность этих формул в том, что многие из используемых в них чисел — 1, 4, 32, 64, 256, а также 2^{4n} и 2^{10n} — являются степенями двойки, что, конечно, сильно упрощает расчеты в двоичной системе, которая используется для внутренних операций в компьютерах. После этого открытия хлынул целый поток новых формул для π и некоторых других интересных чисел. Рекорд вычисления одиночных двоичных цифр числа π обновляется регулярно: в 2010 г. Николас Ши из Yahoo рассчитал двухквадрильонную двоичную цифру π, которой оказался 0.

При помощи тех же формул можно находить отдельные цифры числа π в арифметических операциях с основаниями 4, 8 и 16. Ни для каких других оснований ничего подобного не известно; в частности, мы не можем вычислять десятичные цифры по отдельности. Существуют ли в принципе такие формулы? До открытия формулы Бэйли–Боруэйна–Плаффа никому даже в голову не приходило, что можно это делать хотя бы в двоичной системе.

4

Загадки картографии

Теорема о четырех красках

Многие великие задачи уходят корнями в глубокие и сложные вопросы давних и хорошо известных областей математики. Это те случаи, когда серьезные препятствия вдруг возникают уже после того, как эта область была тщательно изучена. Они, как правило, имеют технический характер, и все заинтересованные лица заранее знают, что они очень сложны, — еще бы, ведь многие специалисты пытались одолеть их и потерпели неудачу. При этом для соответствующей области часто уже разработаны множество мощных методик и объемный математический аппарат, которым может воспользоваться всякий подготовленный человек, но при этом, если задача до сих пор не решена, значит, все очевидные способы воспользоваться этими методиками уже испробованы и *не сработали*. Так что для решения этой задачи нужно либо использовать испытанные инструменты каким-то другим способом, либо изобретать новые.

Бывало и так, и этак.

Но существуют великие задачи, у которых все иначе. Они появляются из ниоткуда — небрежный чертеж на песке, заметка на полях книги, мимолетная причуда. Их формулировки просты, но поскольку вокруг них нет обширного математического фона, то нет и традиционных методов и подходов к ним. Иногда проходит много лет, прежде чем становится ясен уровень сложности задачи: кажется, что должен существовать какой-то хитрый, но несложный трюк, при помощи которого ее можно решить, и что решение не займет и полстранички. Задача о четырех красках относится именно к этой категории. Прошло не одно десятилетие, прежде чем математики начали осознавать, насколько она сложна. Мало того, большую часть этого времени все думали, что она *уже* решена, причем именно на нескольких страничках. Вообще, задача казалась второстепенной, и мало кто принимал ее всерьез, а когда это все же происходило, то в существовавшем вроде бы решении обнаруживались изъяны. Окончательное решение устранило все недостатки, но к тому моменту дискуссия стала настолько сложной, что пришлось привлекать на помощь мощные компьютеры.

В конечном итоге оба типа задач, несмотря на разное происхождение, схожи тем, что решение тех и других невозможно без новых подходов. Несмотря на то что задачи первого типа коренятся в хорошо изученных областях математики, традиционных методов для их решения не хватает. А задачи второго типа не принадлежат ни к одной известной области — более того, стимулируют возникновение новых, — и поэтому традиционных методов, которые можно было бы к ним применить, просто не существует. В обоих случаях решение задачи требует изобретения новых методов и установления новых связей с существующим массивом математических знаний.

Происхождение задачи о четырех красках известно, и оно — не математическое. В 1852 г. молодой южноафриканский математик и ботаник Фрэнсис Гутри, готовившийся к получению ученой степени по юриспруденции, попытался раскрасить графства на карте Англии. Он хотел быть уверенным, что любые два смежных графства можно будет раскрасить в разные цвета, чтобы границы между ними были хорошо различимы. Гутри выяснил, что для выполнения задачи ему будет достаточно четырех различных цветов, и после некоторого количества экспериментов убедил себя в том, что это заявление будет верным для абсолютно любой карты. Говоря о «смежных» графствах, он имел в виду, что эти графства имеют общую границу ненулевой длины; если же два графства соприкасались в точке или, к примеру, в нескольких изолированных точках, их можно было при необходимости раскрасить в один и тот же цвет. Без этой оговорки число цветов может быть бесконечным, поскольку в одной точке может встретиться неограниченное число регионов (см. рис. 8 слева).

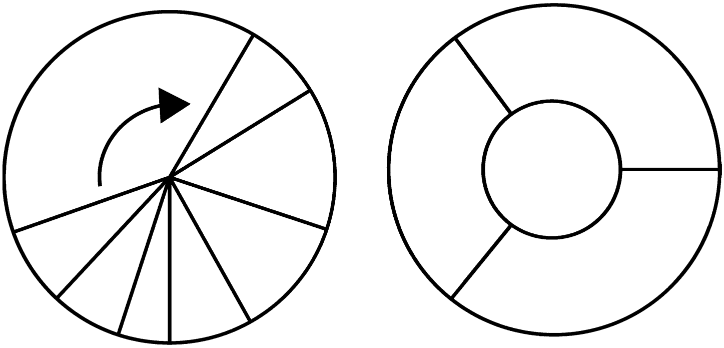


Рис. 8. Любое число регионов может встретиться в точке *(слева)*. Необходимо по крайней мере четыре краски *(справа)*

Заинтересовавшись, не является ли его вывод известной математической теоремой, Гутри задал этот вопрос своему брату Фредерику, изучавшему в то время математику под руководством известного, но эксцентричного ученого Огастеса де Моргана в Университетском колледже Лондона. Де Морган не знал ответа на этот вопрос, поэтому написал еще более известному математику — ирландцу сэру Уильяму Гамильтону:

> «Один мой студент [позже выяснилось, что это был Фредерик Гутри] попросил меня сегодня объяснить один факт, про который мне ничего не было известно, — и я до сих пор не уверен, что это действительно факт. Он говорит, что если некая фигура разделена на части любым спосо-

> бом и ее части раскрашены по-разному, так что фигуры с общей границей в виде *линии* любой длины окрашены в разные цвета, то для этого может потребоваться четыре краски, но не больше… Вопрос: нельзя ли придумать случай для пяти или более красок… Что скажете? И был ли этот факт, если это правда, замечен ранее?»

Фредерик позже упоминал некое «доказательство», предложенное его братом, но говорил также, что основной идеей там был рисунок, примерно соответствующий рис. 8, а он доказывает лишь, что меньше, чем четырьмя красками, не обойдешься.

Ответ Гамильтона был краток: «Я вряд ли займусь в ближайшее время вашим “кватернионом” красок». В то время Гамильтон работал над алгебраической системой, которой суждено было на всю жизнь стать его пунктиком и любимым коньком. Это система, аналогичная комплексным числам, но включающая четыре типа чисел вместо двух (действительные и мнимые) в комплексной системе. Свои числа он называл «кватернионами». Предложенная им система до сих пор сохраняет свое значение в математике. Мало того, сегодня ее роль, вероятно, более важна, чем во времена Гамильтона. Но высот, о которых мечтал автор, она так и не достигла. Гамильтон просто пошутил в академическом стиле, употребив слово «кватернион» по отношению к четырем краскам. Долгое время действительно казалось, что между его системой и задачей о четырех красках нет никакой связи. Однако задачу можно переформулировать так, что она становится утверждением о кватернионах, так что Гамильтон, сам того не желая, попал в яблочко.

Де Морган, потерпев неудачу в поиске доказательства, рассказал о задаче всем своим знакомым математикам в надежде на то, что кто-нибудь сможет предложить полезную идею. В конце 1860-х гг. американский логик, математик и философ Чарльз Пирс заявил, что нашел решение задачи о четырех красках, а также ответы на аналогичные вопросы о картах на более сложных поверхностях. Предполагаемое доказательство так и не было опубликовано, но вряд ли доступные ему методы были адекватны задаче.

Хотя в задаче о четырех красках говорится вроде бы о картах, сама она не имеет применения в картографии. Практика раскраски карт отражает в основном политические различия, и если при этом соседние регионы должны иметь один цвет, то их и красят одинаково. Смысл этой задачи лежит в области чистой математики — новой области, которая тогда только начала развиваться — топологии. Это «геометрия на резиновом листе», в которой фигуры можно непрерывно деформировать любым способом. Но даже там задача о четырех красках не укладывалась в основное русло исследований, а представлялась всего лишь диковинкой.

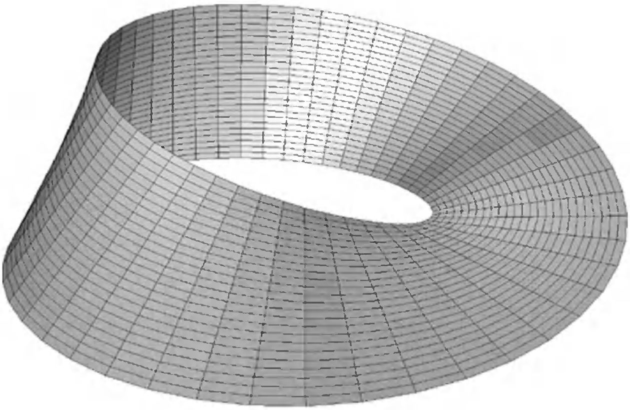


Рис. 9. У ленты Мёбиуса всего одна сторона

Одним из пионеров топологии был Август Мёбиус, известный сегодня благодаря своей односторонней ленте (см. рис. 9). Модель такой ленты несложно изготовить: для этого нужно взять полоску бумаги, свернуть ее в кольцо, похожее на короткий толстый цилиндр, повернуть один из концов на 180° и склеить концы. Однажды друг Мёбиуса лингвист Бенджамин

Вейске загадал ему загадку: может ли индийский царь разделить свое царство на пятерых сыновей так, чтобы часть, принадлежащая одному принцу, имела границу ненулевой длины с частями всех остальных? Мёбиус задал эту загадку своим студентам в качестве упражнения, но на следующей лекции извинился за то, что попросил их сделать невозможное. Подразумевалось, что он может *доказать* невозможность ее решения[16].

Эту загадку трудно представить геометрически, поскольку формы отдельных частей могут, в принципе, быть очень сложными. Для успешного продвижения в решении этой задачи следует ввести серьезное упрощение: сказать, что существенно только то, какие регионы граничат и как их общие границы расположены относительно друг друга. Эта топологическая информация не зависит от конкретных форм и может быть представлена в четкой и простой форме, известной как граф, или в наши дни — сеть (это более выразительный термин).

Сеть — чрезвычайно простое понятие: набор вершин (они обозначаются точками), некоторые из которых связаны ребрами (обозначаются линиями). Возьмем произвольную карту (см. рис. 10 слева). Чтобы представить ее в виде сети, поставим в каждой области по точке (см. рис. 10 в середине). Там, где две области имеют общий участок границы, соединим соответствующие точки линией, проходящей через этот участок. Если две области имеют несколько общих участков границы, проведем через каждый по отдельной линии. Проделаем все это для всех областей и всех участков границы так, чтобы линии не пересекались друг с другом и не имели самопересечений, а встречались только в точках. Затем выбросим первоначальную карту и оставим себе только точки и линии. Это двойственная сеть — двойник нашей карты (см. рис. 10 справа).

Слово «двойственный» используется потому, что при этой процедуре области, линии и точки (пересечения областей) превращаются в точки, линии и области. Область на карте соответствует точке двойственной сети. Участок границы на карте соответствует линии двойственной сети; не той же самой

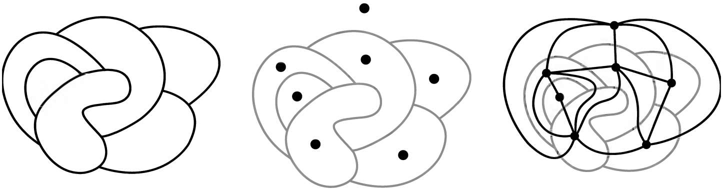


Рис. 10. Карта *(слева)*. Поставим по точке в каждой области *(в середине)*. Соединим точки через границы, чтобы сформировать двойственную сеть (в нее входят только черные точки и линии) *(справа)*

линии, а линии, которая пересекает границу и связывает соответствующие точки. Точка, в которой на карте сходятся три или больше областей, соответствует области двойственной сети, ограниченной со всех сторон линиями. Так что двойственная сеть — сама по себе карта, поскольку линии здесь ограничивают области; кроме того, оказывается, что двойственной схемой к двойственной схеме является первоначальная карта плюс-минус кое-какие технические подробности, исключающие ненужные точки и линии.

Рассматривая двойственную сеть, задачу о пяти принцах можно сформулировать иначе: можно ли соединить пять точек на плоскости непересекающимися линиями? Ответ — нет, а ключ к нему — формула Эйлера, согласно которой, если карта на плоскости состоит из F участков (областей), E ребер (линий) и V узлов (точек), то $F + V - E = 2$. Здесь остальная плоскость, оставшаяся вне сети, считается одной большой областью. Эта формула в свое время стала первым указанием на то, что топологические вопросы достойны рассмотрения. Она вновь появится в главе 10.

Доказательство того, что задача о пяти принцах не имеет решения, начинается с предположения о том, что такое решение существует, и это приводит к противоречию. Любое решение должно иметь число точек $V = 5$. Поскольку каждая пара точек соединяется линиями, а точек у нас 10 пар, то $E = 10$. Тогда по теореме Эйлера $F = E - V + 2 = 7$. Области двойственной сети ограничены замкнутыми петлями линий, и каждую пару точек соединяет только одна линия, поэтому каждая из петель должна содержать по крайней мере три линии. Если областей семь, то линий получается по крайней мере 21… Правда, каждая из них считается дважды, поскольку разделяет две области. Так что линий по крайней мере 10,5. Число линий должно быть целым, значит, на практике у нас должно быть по крайней мере 11 линий. Однако мы уже знаем, что линий у нас 10. Это логическое противоречие доказывает, что такой сети не существует. Царь не сможет разделить свои земли так, как ему хочется.

В подобных рассуждениях обнадеживает то, что элегантные топологические методы позволяют нам доказывать интересные и неожиданные факты о картах. Однако, вопреки распространенному мнению, которое де Морган, судя по всему, разделял, невозможность решения задачи о пяти индийских принцах *не доказывает* теорему о четырех красках. Доказательство может быть неверным, даже если само умозаключение верно, или по крайней мере никому не известно о его неверности. Если где-то в предполагаемом доказательстве мне встретится треугольник с четырьмя сторонами, я прекращу читать, поскольку это доказательство неверно. При этом не имеет значения, что происходит в нем позже или какой из этого делается вывод. Наш ответ на загадку индийских принцев показывает лишь, что один из способов опровержения теоремы о четырех красках не работает. Однако из этого не следует, что не может работать какой-нибудь *другой* способ. В принципе, может существовать множество причин, по которым карту не удастся раскрасить в четыре цвета. Существование пяти областей, каждая из которых граничит со всеми остальными, лишь одна из этих потенциальных причин. Пока не доказано обратное, может существовать очень сложная карта, скажем, из 703 регионов, на которой, даже если вам удастся раскрасить в четыре цвета 702 из них, последний оставшийся все равно потребует пятой краски. Конечно, этот регион должен будет граничить по крайней мере с четырьмя другими, но это вполне представимо и не требует выполнения условий задачи о пяти принцах. Если бы подобная карта нашлась, она доказала бы, что четырех красок недостаточно. Любое доказательство должно исключить все подобные препятствия. И это утверждение сохраняет силу даже в том случае, если я не смогу продемонстрировать вам конкретный пример такой карты.

На какое-то время задача о четырех красках полностью пропала из виду, но в 1878 г. Артур Кейли упомянул ее на заседании Лондонского математического общества, и она вновь вызвала интерес. Несмотря на название, Общество это представляло всю британскую (или по крайней мере всю английскую) математику, а его основателем был де Морган. Кейли поинтересовался, удалось ли кому-нибудь получить решение этой задачи, и вскоре после заседания его вопрос был опубликован в журнале *Nature*. Годом позже он опубликовал обширную статью на эту тему в «Трудах Королевского географического общества»[17]. Поскольку речь в задаче вроде бы шла о картах, издание показалось подходящим для публикации. Может быть, статью автору даже заказали. Но на самом деле выбор оказался не слишком удачным — ведь картографам решение этой задачи не нужно и не интересно, разве что из чистого любопытства. По той же причине, к несчастью, мало кто из математиков заметил эту статью, а жаль: в ней Кейли объяснил, почему задача может оказаться сложной.

В главе 1 я писал, что доказательство чем-то напоминает сражение. В военном деле четко различаются тактика и стратегия. Тактика — это искусство выигрывать локальные сражения, а стратегия определяет общую структуру кампании. Тактика определяет передвижение каждой войсковой части; стратегия формирует обширные планы, в рамках которых на каждой стадии возможны самые разные тактические решения. Статья Кейли не блистала тактическими находками, но содержала легчайший намек на стратегию, которая по прошествии времени позволила расколоть этот орешек и решить задачу о четырех красках. Кейли заметил, что добавление областей последовательно, по одной, ничего не дает, если следовать очевидному ходу рассуждений. Но, может быть, если найти другой, менее очевидный ход, из этого что-нибудь получится.

Предположим, мы возьмем произвольную карту и уберем оттуда одну область —сольем ее с соседней или сожмем в точку. Предположим также, что получившуюся карту можно раскрасить в четыре цвета, и мы так и сделаем. А теперь вернем удаленную область на место. Если нам повезет, ее соседи окажутся раскрашенными только в три цвета. Тогда нам останется всего лишь закрасить восстановленную область четвертым — свободным — цветом. Кейли указал, что эта процедура может и не сработать, поскольку соседи нашей области могут оказаться раскрашенными в четыре разных цвета. Но это не означает, что все плохо. Такое препятствие можно обойти двумя способами: сделать вывод либо о том, что мы выбрали не ту область, либо о том, что мы неверно раскрасили уменьшенную карту.

Действуя на основании ничем не подтвержденных предположений (это очень эффективный способ формирования рабочих идей, хотя в какой-то момент их все равно придется обосновать), считаем, что подобные препятствия всегда устранимы. Тогда получается, что любую карту можно раскрасить в четыре цвета, если известно, что некую меньшую карту можно так раскрасить. Может показаться, что такой вывод ничего нам не дает: как мы узнаем, что какую-то меньшую карту можно раскрасить нужным образом? Ответ в том, что эту же процедуру можно применить к меньшей карте, а затем к еще меньшей карте… и т. д. В конце концов, мы доберемся до карты настолько маленькой, что в ней будет всего четыре области, и это гарантирует, что ее можно раскрасить в четыре цвета. Теперь пройдем тот же путь в обратном направлении, на каждом шагу раскрашивая карту чуть побольше, чем в прошлый раз, и, в конце концов, доберемся до первоначальной карты.

Подобные рассуждения называют «доказательством по индукции». Это стандартный метод доказательства наиболее формализованных формулировок, и логику, на которой он основан, можно сделать строгой. Предложенная Кейли стратегия доказательства становится более понятной, если переформулировать ее с использованием логически эквивалентной концепции минимального контрпримера. В данном контексте контрпримером можно считать любую гипотетическую карту, которую невозможно раскрасить в четыре краски. Такая карта будет минимальной, если любую меньшую карту (т. е. карту с меньшим числом областей) *можно* раскрасить нужным образом. Если хотя бы один контрпример существует, то должен существовать и минимальный контрпример: чтобы его найти, нужно просто взять контрпример с минимальным возможным числом областей. Поэтому если минимального контрпримера не существует, то контрпримеров не существует вообще. А если их нет, то теорема о четырех красках верна.

Доказательство по индукции сводится примерно к следующему. Предположим, мы можем доказать, что минимальный контрпример всегда можно раскрасить в четыре краски, если можно раскрасить так некую связанную с ним меньшую карту. Тогда минимальный контрпример не может считаться собственно контрпримером. Поскольку эта карта минимальна, *все* меньшие карты можно раскрасить в четыре краски, поэтому, исходя из утверждения, которое, согласно принятому нами предположению, может быть доказано, то же верно в отношении первоначальной карты. Следовательно, минимального контрпримера не существует, а значит, не существует контрпримеров вообще. Эта идея сдвигает фокус проблемы, позволяя рассматривать не все карты сразу, а только гипотетические минимальные контрпримеры, и определяет процедуру редукции — способ последовательно вывести четырехкрасочность первоначальной карты из четырехкрасочности некой соответствующей меньшей карты.

Но зачем возиться с минимальными контрпримерами, не лучше ли поискать обычные? Это вопрос методики. Хотя первоначально мы не знаем, существуют ли контрпримеры, одно из парадоксальных, но очень полезных свойств этой стратегии заключается в том, что мы можем многое сказать о том, как должны выглядеть именно минимальные контрпримеры, если они существуют.

Для этого необходима способность рассуждать логически о гипотетических вещах — жизненно важное умение для любого математика. Чтобы дать вам почувствовать характер процесса, я докажу теорему о *шести* красках. Для этого мы позаимствуем прием из загадки о пяти принцах и переформулируем все в терминах двойственной сети, в которой области становятся точками. В этом случае задача о четырех красках эквивалентна другому вопросу: если на плоскости задана сеть, линии которой не пересекаются, можно ли раскрасить в четыре цвета *точки* так, чтобы две точки, соединенные линией, всегда были разного цвета? Точно так же можно переформулировать задачу с любым количеством красок.

Чтобы проиллюстрировать мощь метода минимальных контрпримеров, я докажу с их помощью, что любую плоскую сеть можно раскрасить в шесть цветов. Здесь главным нашим инструментом вновь станет формула Эйлера. Для точки плоской двойственной сети соседними точками назовем те, что соединены с ней линиями. У каждой точки может быть и множество соседей, и всего несколько. Можно показать, что, в соответствии с формулой Эйлера, у некоторых точек должно быть мало соседей. Точнее говоря, в плоской сети все точки не могут иметь по шесть и больше соседей. Доказательство этого момента я поместил в примечание, чтобы не прерывать ход мысли[18]. Этот факт дает нам рычаг, необходимый для того, чтобы разбить задачу на более мелкие подзадачи. Рассмотрим гипотетический минимальный контрпример для теоремы о шести красках. Это сеть, которую невозможно раскрасить в шесть разных цветов, притом что любую меньшую сеть так раскрасить можно. А теперь я доказываю, что такая карта не может существовать. Согласно приведенному выше следствию из формулы Эйлера, в ней должна быть хотя бы одна точка, у которой пять или меньше соседей. Временно уберем эту точку и линии, соединяющие ее с соседями. В получившейся сети меньше точек, поэтому, исходя из минимальности контрпримера, ее можно раскрасить в шесть цветов. (Этот шаг, кстати говоря, мы не сможем сделать, если наш контрпример будет не минимальным.) А теперь вернем удаленную точку и ее линии на место. Эта точка имеет не более пяти соседей, так что шестой цвет для нее всегда найдется. Покрасим ее — и получим успешно раскрашенный в шесть цветов минимальный контрпример; но тогда получается, что это никакой не контрпример. Значит, минимальных контрпримеров для теоремы о шести красках не существует, а значит, теорема верна.

Это внушает оптимизм. До сих пор, насколько нам известно, для раскраски некоторых карт могло потребоваться 20 цветов, или 703, или несколько миллионов. Теперь мы знаем, что такие карты не более реальны, чем горшок золота под концом радуги. Мы знаем, что конкретного ограниченного числа красок точно хватит на *любую* карту. Это настоящий триумф метода минимальных контрпримеров. Математики, посмотрев на него, взялись за дело с еще большим энтузиазмом, надеясь усилить аргументацию и постепенно заменить шесть красок на пять, а если повезет, и на четыре.

Юристы иногда тоже интересуются математическими задачами. Адвокат по имени Альфред Кемпе присутствовал на том заседании, где Кейли упомянул задачу о четырех красках. В свое время он под руководством Кейли изучал математику в Кембридже, и за годы его интерес к этой науке нисколько не ослаб. Не прошло и года после заседания, а Кемпе уже убедил себя, что ему удалось справиться с задачей. Свое решение он опубликовал в 1879 г. в недавно основанном журнале *American Journal of Mathematics*. Еще через год он опубликовал упрощенное доказательство, где были исправлены некоторые ошибки, присутствовавшие в первом. Вот как он подошел к вопросу: «Очень небольшое изменение в части карты может привести к необходимости перекрашивать ее целиком. В результате достаточно сложного поиска мне удалось отыскать слабое звено, которое позволило одержать победу».

Я изложу идеи Кемпе в терминах двойственной сети. Опять же он начал с формулы Эйлера и следующего из нее вывода о существовании точки с тремя, четырьмя или пятью соседями. (Точка с двумя соседями лежит на линии и никак не влияет на структуру сети или карты: на нее можно просто не обращать внимания.)

Если существует точка с тремя соседями, то процедуру, которую я использовал для доказательства теоремы о шести красках, можно применить и к четырехкрасочному варианту. Удаляем саму точку и линии, которые в ней сходятся, раскрашиваем в четыре краски результат, возвращаем точку и линии на место и окрашиваем ее в оставшийся свободным цвет. Поэтому мы можем считать, что точки с тремя соседями не существует.

Если существует точка с четырьмя соседями, то вышеописанная методика не срабатывает, потому что при возвращении точки свободного цвета может и не оказаться. Кемпе придумал хитрый способ обойти это препятствие: он предложил так же точно удалить точку, но после этого поменять расцветку получившейся меньшей карты так, чтобы два из четырех ее бывших соседей получили один и тот же цвет. После такой модификации у соседей удаленной точки окажется не больше трех цветов — и в нашем распоряжении окажется свободный четвертый. Основная идея перекраски схемы по Кемпе заключается в том, что две соседние точки должны быть разных цветов — скажем, синего и красного, а еще в схеме используются зеленый и желтый. Если обе оставшиеся точки окажутся зелеными или желтыми, то второй цвет окажется свободным и может быть использован для удаленной точки. Исходя из этого, считаем, что одна из них зеленая, а вторая — желтая. Теперь найдем все точки, которые соединены с синей точкой последовательностью линий, проходящих только через синие и красные точки, и назовем их красно-синей цепочкой Кемпе[19].

По определению, любой сосед любой точки в цепочке Кемпе, не принадлежащий цепочке, должен быть зеленым или желтым, поскольку синий или красный сосед там уже есть. Обратите внимание, что замена цветов в пределах цепочки Кемпе (синий на красный, и наоборот) дает новый вариант карты, в которой по-прежнему выполняется ключевое условие о том, что соседние точки должны быть разных цветов (см. рис. 11).

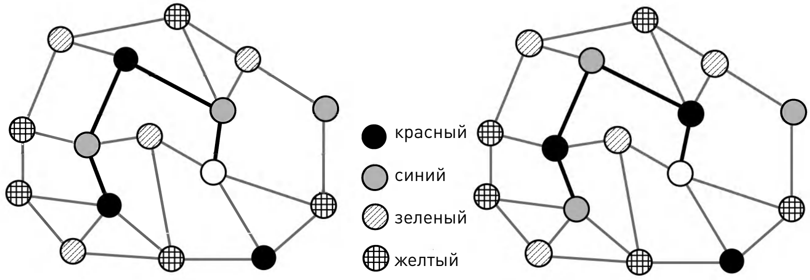


Рис. 11. Замена цветов в цепочке Кемпе (жирные черные линии), связанной с точкой четвертого порядка (белой), которая имеет соседей всех четырех цветов: *слева:* первоначальные цвета; *справа:* если цвета в цепочке поменять, синий останется свободным и может быть использован для окраски белой точки

Если красный сосед нашей точки не является частью выделенной сине-красной цепочки, проведите такую замену. Синий сосед точки сделается красным, а красный останется красным по-прежнему. Теперь соседи нашей точки окрашены не более чем в три цвета: красный, зеленый и желтый, что позволяет нам окрасить точку в синий цвет — и дело сделано. Однако сине-красная цепочка может описать петлю и замкнуться на синем соседе нашей точки. Если так, оставьте в покое синий и красный цвета и проделайте ту же операцию с ее желтыми и зелеными соседями. Начните с зеленой точки и сформируйте желто-зеленую цепочку Кемпе. Заметьте: она не сможет замкнуться на желтого соседа, поскольку на ее пути непременно встретится предыдущая красно-синяя цепочка. Поменяйте желтый и зеленый цвета в цепочке местами, и дело сделано.

Остается последний случай: когда не существует точек с тремя или четырьмя соседями, но по крайней мере одна точка имеет пять соседей. Кемпе предложил аналогичное, но более сложное правило перекраски точек, которое, на первый взгляд, успешно решало и эту проблему. Вывод: теорема о четырех красках верна, и доказал ее Кемпе. Эта заявление попало даже в средства массовой информации: американский журнал *The Nation* упомянул решение Кемпе в своем обзоре.

Казалось, с проблемой четырех красок было покончено, и математики в большинстве своем с этим согласились. Правда, Питер Тэт продолжал поиски более простого решения и время от времени публиковал статьи на эту тему. Исследования привели его к нескольким полезным открытиям, но простое доказательство по-прежнему не давалось.

И тут на сцене появляется преподаватель математики из Университета Дарема Перси Хивуд, прозванный за свои великолепные ухоженные усы «Котом». Еще студентом в Оксфорде он услышал от профессора геометрии Генри Смита о теореме о четырех красках. Смит сказал ему, что теорема эта, хотя, вероятно, и верна, но не доказана, так что у Хивуда есть шанс. Кроме того, как-то он наткнулся на статью Кемпе и попытался ее понять. Результат своих размышлений Хивуд опубликовал в 1889 г. под названием «Теорема о раскраске карты», высказав при этом сожаление, что цель его статьи более «деструктивна, чем конструктивна, ибо в ней будет показано, что в признанном, кажется, на сегодня доказательстве есть дефект». Кемпе допустил ошибку.

Ошибка была достаточно тонкой и возникала в схеме перекраски в том случае, когда у удаляемой точки было пять соседей. В некоторых случаях изменение цвета одной точки (по схеме Кемпе) могло повлечь за собой невозможность дальнейших изменений. При этом Кемпе считал, что если какая-то точка меняет цвет, то происходит это лишь один раз. Хивуд же нашел карту (или сеть), в которой схема перекраски по Кемпе не срабатывала, и тем самым опроверг его доказательство. Кемпе, узнав об этом, без промедления признал ошибку и добавил, что ему «не удалось исправить этот дефект». Теорема о четырех красках вновь ждала желающих помериться с ней силой.

Хивуд отыскал в этой истории небольшое утешение для Кемпе: его метод успешно доказывал теорему о пяти красках. Кроме того, Хивуд работал еще над двумя обобщенными вариантами задачи: над вариантом с империями, где области могли состоять из нескольких несвязанных кусков, которые все требовали одного цвета, и над картами на более сложных поверхностях. Аналогичная задача на сфере решается точно так же, как на плоскости. Представьте себе карту на сфере, причем разверните сферу так, чтобы Северный полюс оказался внутри одной из областей. Теперь, если удалить точку полюса, то сферу с отверстием можно развернуть в поверхность, топологически эквивалентную бесконечной плоскости. Регион, в котором находился полюс, развернется в бесконечное пространство, окружающее карту. Однако, помимо сферы, существуют и другие, более интересные поверхности. Среди них тор, напоминающий по форме бублик с дыркой (см. рис. 12 слева).

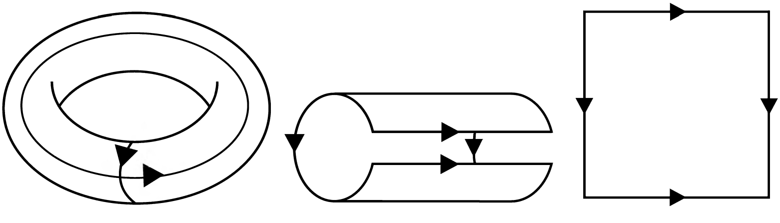


Рис. 12. Тор можно вскрыть и развернуть в квадрат

Существует полезный способ визуализации тора, часто упрощающий математикам жизнь. Если разрезать тор вдоль двух замкнутых кривых (см. рис. 12 в середине), то можно развернуть его поверхность так, чтобы получился квадрат (см. рис. 12 справа). Такая трансформация меняет топологию тора, но эту сложность можно обойти, если объявить противоположные стороны получившегося квадрата тождественными. В результате (а строгое определение позволяет точно сформулировать принцип) мы договариваемся считать, что соответствующие точки на этих сторонах совпадают. Чтобы представить, почему так, посмотрите на рисунки в обратном порядке. Мы скатываем квадрат в трубочку, и противоположные его стороны действительно склеиваются, затем сгибаем трубочку в кольцо и соединяем концы. Готово. А теперь самое интересное. Не обязательно на самом деле скручивать квадрат в трубочку и соединять соответствующие стороны. Можно работать с плоским квадратом, достаточно просто помнить о том, что его противоположные стороны — это одно и то же. Всему, что мы будем делать на торе, включая и рисование кривых, имеется точное соответствие на квадрате. Хивуд доказал, что для раскрашивания любой карты на торе необходимо и достаточно семи красок. Рис. 13 (слева) показывает, что семь цветов необходимо; при этом квадрат, как уже говорилось, представляет поверхность тора. Обратите внимание, как сходятся участки на противоположных сторонах квадрата. Существуют поверхности, подобные тору, но имеющие больше отверстий (см. пример на рис. 13 справа). Число отверстий в такой фигуре называется родом и обозначается буквой g (*genus* — род). Хивуд придумал формулу для числа красок, необходимых для карты на торе с g отверстиями, если $g \geq 1$: это наибольшее целое число, меньшее или равное

$$\frac{7+\quad 48g+1}{2}.$$

При g от 1 до 10 формула выдает следующие результаты:

7 8 9 10 11 12 12 13 13 14.

Число красок, определяемое формулой, растет медленнее, чем род тора, и нередко добавление лишнего отверстия в торе ничего не меняет. Это удивительно, потому что каждое допол-

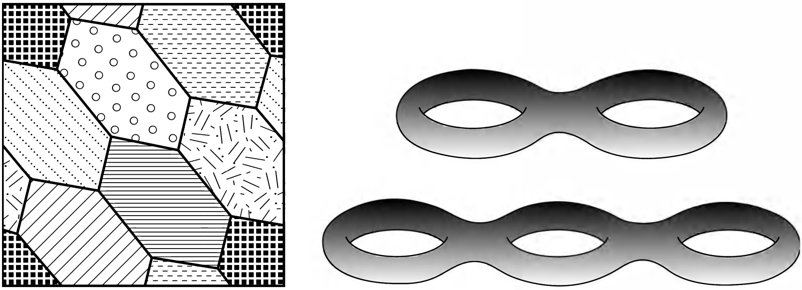


Рис. 13. *Слева*: семицветная карта на торе. Тор представлен как квадрат, противоположные стороны которого можно свернуть и склеить друг с другом. *Справа*: торы с двумя и тремя отверстиями

нительное отверстие дает бóльшую свободу для изобретения сложных карт.

Хивуд не просто извлек эту формулу из воздуха. Она возникла из обобщения метода, при помощи которого я доказывал теорему о шести красках на плоскости. Он сумел доказать, что такого числа красок всегда достаточно. Однако вопрос о том, нельзя ли сделать это число меньше, оставался открытым еще много лет. Примеры для небольшого значения рода показывали, что оценка Хивуда — наилучшая из возможных. Только в 1968 г. Герхардт Рингель и Джон Янгс заполнили остававшиеся пробелы и доказали на базе собственных и чужих работ, что формула верна. Они использовали при этом комбинаторные методы, основанные на сетях особого рода и достаточно сложные, чтобы заполнить собой целую книгу.

При $g = 0$, т. е. для карт на сфере, формула Хивуда дает четыре краски, но его доказательство достаточности на сфере не работает. Несмотря на значительный прогресс для поверхностей хотя бы с одним отверстием, первоначальная теорема о четырех красках никуда не делась. Немногочисленные математики, которые готовы были бросить свои силы на решение этого вопроса, настроились, говоря языком военных, на длительную осаду. Задача оказалась неприступной крепостью, но желающие завоевать ее надеялись построить еще более мощные осадные машины и понемногу, по кусочку, разбить и обрушить стены. Машины были построены, но стены продолжали стоять. Однако атакующие постепенно все больше узнавали о том, как не следует решать эту задачу, и о препятствиях, возникающих на этом пути. Таким образом неудачи создали почву для появления новой стратегии. Она стала естественным продолжением методов Кемпе и Хивуда и состоит из трех частей. Я перечислю их, используя понятия двойственных сетей, поскольку на сегодня это стандартный подход.

1. Рассмотреть минимальный контрпример.
2. Составить список неустранимых конфигураций — меньших сетей, таких, что любой минимальный контрпример обязательно должен содержать какую-нибудь из них.
3. Доказать, что каждая из неустранимых конфигураций сократима. Иными словами, если меньшая сеть, полученная при удалении неизбежной конфигурации, может быть раскрашена в четыре цвета, то эти цвета можно перераспределить таким образом, что при возвращении неустранимой конфигурации раскраску в четыре цвета можно распространить и на нее тоже.

Объединив эти три шага, мы можем доказать, что минимального контрпримера не существует. Если бы он существовал, то обязательно содержал бы хотя бы одну из неустранимых конфигураций. Но остальная часть сети меньше по размеру, поэтому из минимальности контрпримера следует, что он может быть раскрашен в четыре цвета. А сводимость подразумевает, что исходная сеть тоже может быть раскрашена в четыре цвета. Это противоречие.

Исходя из этих посылок, Кемпе составил (причем совершенно верно) список неустранимых конфигураций: это точка с выходящими из нее тремя, четырьмя или пятью линиями (см. рис. 14). Кроме того, Кемпе корректно доказал, что первые две конфигурации сводимы, однако ошибся с доказательством сводимости третьей конфигурации. На самом деле она несводима. Отсюда предложение: замените эту нехорошую конфигурацию более длинным набором конфигураций, следя за тем, чтобы полный набор оставался неизбежным. Проделайте это таким образом, чтобы каждая конфигурация в наборе была сводимой. Иными словами, найдите неустранимой множество сводимых конфигураций. Если у вас получится, это будет означать, что вы доказали теорему о четырех красках.

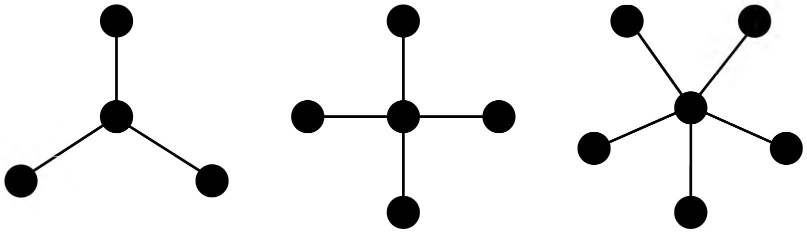


Рис. 14. Набор неустранимых конфигураций Кемпе

Такого набора, вообще говоря, может и не быть в природе, но стратегия сама по себе заслуживает внимания, тем более что ничего лучшего никто предложить не сумел. Правда, у этого метода есть один недостаток. С одной стороны, чем длиннее список конфигураций, тем больше шансов на то, что они действительно неизбежны, а это хорошо. С другой стороны, чем длиннее список, тем меньше вероятность того, что каждая конфигурация в нем окажется сводимой; а если это не так, то все доказательство рушится. Эта опасность становится тем острее, чем больше в списке конфигураций, а это плохо. С *третьей* стороны, более длинный список предоставляет больше возможностей для выбора сводимых конфигураций, и это хорошо. С четвертой — он увеличивает объем работы, необходимый для доказательства сводимости, и это плохо. А с пятой стороны, хороших способов сделать это просто не существовало.

Именно такие препятствия делают великие задачи великими.

Итак, в течение какого-то времени события развивались так: осаждающим удавалось иногда отбить от стены камешек, но это никак не сказывалось на ней. При этом весь остальной математический мир смотрел на эту осаду позевывая, если вообще обращал на нее внимание. Но кое-кто — его звали Генрих Хееш — уже сооружал более мощный таран. Его вкладом в решение задачи стал метод доказательства сводимости конфигурации, который автор называл «разрядкой». По его мысли, точки в сети следовало рассматривать как приблизительный аналог электрических зарядов, а раскрашивание — как перетекание электричества от одной точки к другой.

Но даже при помощи метода Хееша вручную искать неизбежный набор сводимых конфигураций было невероятно сложно. Отдельные конфигурации при этом, вероятно, были бы достаточно небольшими, но их количество… Хееш продолжал упорно работать, а в 1948 г. даже прочитал курс лекций на эту тему. Он полагал, что полный набор конфигураций должен включать порядка 10 000 штук. На тот момент он успел доказать сводимость 500 комбинаций. На одной из лекций Хееша присутствовал молодой человек по имени Вольфганг Хакен. Позже он признавался, что мало что понял из того, о чем говорил Хееш, но некоторые его рассуждения Хакену запомнились. Он продолжил изучать топологию и позже совершил крупное открытие в теории узлов. Это побудило его взяться за гипотезу Пуанкаре (см. главу 10). Исследуя один из подходов к проблеме, Хакен разложил все возможные случаи на 200 вариантов, решил 198 из них и еще 13 лет безуспешно сражался с двумя оставшимися. После этого он сдался и перешел к задаче о четырех красках. Очевидно, Хакен любил по-настоящему сложные проблемы, но его беспокоила мысль о том, что с 10 000 комбинаций Хееша может произойти нечто подобное. Представьте только: успешно разобраться с 9998 комбинациями и застрять на двух последних. Поэтому в 1967 г. он пригласил Хееша к себе в Университет штата Иллинойс, чтобы спросить совета.

В те дни компьютеры уже начинали потихоньку проникать в мир математики, но тогда они были громадными машинами, которые занимали целые здания, а не стояли спокойно на столе и не умещались в портфеле. Хакена интересовало, можно ли прибегнуть для решения задачи к помощи компьютеров. Оказалось, что такая мысль уже приходила Хеешу в голову, и он даже сделал примерную оценку сложности этой задачи. Из нее следовало, что даже лучший компьютер, к которому он мог бы получить доступ, с ней не справится. В Иллинойсе, однако, был гораздо более мощный компьютер ILLIAC-IV, и Хакен подал заявку на машинное время. Оказалось, однако, что компьютер еще не готов, и ему посоветовали обратиться в Брукхейвенскую лабораторию на Лонг-Айленде, где имелся Cray 6600. Директором компьютерного центра лаборатории был Ёсио Симамото, тоже очарованный задачей о четырех красках. Хееш и Хакен вытянули счастливый билетик — и получили вожделенный доступ к машине.

Компьютер оправдал ожидания, но Хакен начал подозревать, что его можно использовать намного эффективнее. Они генерировали множество сводимых комбинаций и надеялись собрать когда-нибудь из них полный неизбежный набор, но при такой стратегии много времени напрасно растрачивалось на комбинации, которые в конечном итоге оказывались несводимыми. Может быть, лучше поступить наоборот: сделать неизбежность основной целью, а со сводимостью разобраться позже? Конечно, все равно придется брать комбинации с хорошими шансами на сводимость, но сам по себе этот способ казался более перспективным. Однако к этому моменту Cray в Брукхейвене уже был занят более важными вещами. Хуже того, уже несколько специалистов сказали Хакену, что методы, которыми он хочет воспользоваться, вообще невозможно воплотить в компьютерных программах. Он поверил специалистам и прочел лекцию о том, что задача о четырех красках не может быть решена без помощи компьютеров, но теперь, похоже, получается, что с компьютерами ее тоже не решишь. В общем, Хакен решил оставить попытки.

Среди слушателей на лекции присутствовал и опытный программист Кеннет Аппель, который сообщил Хакену, что эксперты, на которых тот ссылается, вероятно, просто хотели избавиться от него, поскольку на создание подобных программ пришлось бы затратить много усилий при непредсказуемом результате. Аппель считал, что математических задач, которые невозможно запрограммировать, не существует. Вопрос только в том, сможет ли программа получить результат за разумное время. Хакен и Аппель объединили усилия. Стратегия, разработанная как модификация все того же метода разрядки, заставляла вносить изменения в программу, а изменения в программе, в свою очередь, заставляли вносить новые изменения в метод. Этот процесс привел к новой концепции «географически подходящих» конфигураций, которые не содержали определенных неподходящих конфигураций, препятствующих сводимости. Шанс на то, что такая конфигурация окажется сводимой, был заметно выше обычного, а определяющее свойство было несложным и легко поддавалось проверке. Аппель и Хакен решили доказать теоретически, а не на компьютере, что существует неустранимое множество географически подходящих конфигураций. К 1974 г. им это удалось.

Это внушало оптимизм, но ученые понимали, что теперь, скорее всего, произойдет: некоторые из географически подходящих конфигураций непременно окажутся несводимыми, так что придется их исключать и заменять на еще более длинный и сложный набор конфигураций. Программа будет «гоняться за собственным хвостом», и успех будет достигнут только в том случае, если этот хвост удастся догнать. Не желая тратить годы на бесплодные поиски, Хакен и Аппель прикинули, сколько времени может занять процесс. Результаты обнадеживали, поэтому работа была продолжена. Теория и расчеты подпитывали и модифицировали друг друга. Временами компьютер, казалось, начинал жить собственной жизнью и проявлять интеллект, «открывая» полезные свойства конфигураций. Затем администрация университета приобрела новый, очень мощный компьютер — более мощный, чем те, что были доступны на тот момент университетским ученым. После многочисленных протестов и споров половина машинного времени была выделена на научные нужды. Вечно меняющийся список неизбежных конфигураций Аппеля и Хакена стабилизировался на уровне примерно 2000 штук. В июне 1976 г. компьютер выполнил последний тест на сводимость, и доказательство было завершено. Благодаря *The Times* эта история попала в средства массовой информации и стремительно разлетелась по всему миру.

Аппелю и Хакену еще нужно было убедиться, что в доказательстве нет глупых ошибок и упущений, а несколько групп ученых уже устремились по их следам. К июлю, уверившись в действенности своего метода, Аппель и Хакен официально объявили математическому сообществу, что им удалось доказать теорему о четырех красках. Они выпустили препринт — предварительный вариант статьи, который печатается до выхода в свет основной публикации. В то время на публикацию серьезной математической статьи обычно уходило от одного до двух лет. Чтобы не сдерживать прогресс, математикам приходилось искать более быстрые способы познакомить профессиональное сообщество с важными результатами, и препринты были одним из них. В наши дни препринты, как правило, публикуются в Интернете. Полная официальная публикация требует рецензирования, и препринты помогают в ее подготовке — ведь кто угодно может читать их, искать ошибки и сообщать о них авторам, а также предлагать улучшения. Именно поэтому опубликованная версия статьи часто сильно отличается от препринтной.

Окончательное доказательство потребовало около 1000 часов компьютерного времени и содержало 487 правил разрядки. Результаты были опубликованы в двух статьях с 450-станичным приложением, в котором показаны все 1482 конфигурации. На тот момент это был верх совершенства.

Однако основной реакцией математического сообщества стало легкое разочарование. Не результатом как таковым и не замечательным компьютерным достижением. Разочарование вызвал метод. В 1970-е гг. математические доказательства составлялись — и проверялись — вручную. Как я уже говорил в главе 1, доказательство — это рассказ, сюжет которого убеждает вас в истинности того или иного утверждения. Но у этого рассказа не было сюжета. Или если и был, то с большой прорехой на самом интересном месте:

> «Жила-была на свете красивая Гипотеза. Мать говорила ей никогда не заходить в темный опасный лес. Но однажды маленькая Гипотеза-о-Четырех-Красках улизнула из дома и забрела-таки туда. Она знала, что если каждая конфигурация в лесу сводима, то она сможет получить доказательство и стать маленькой Теоремой-о-Четырех-Красках; тогда ее опубликуют в журнале, которым заведует Принц Цвет. Глубоко-глубоко в лесу набрела маленькая Гипотеза на компьютер в шоколаде, внутри которого сидел Волк, притворившийся программистом. И Волк сказал: “Да, они все сводимы”, — и все они жили счастливо и умерли в один день».

Нет, так не годится. Я, конечно, шучу, но прореха в сюжете этой сказки примерно соответствует прорехе в доказательстве Аппеля–Хакена — или по крайней мере тому, что математики в большинстве своем восприняли как прореху в доказательстве. *Откуда нам знать, что Волк сказал правду?*

Мы можем запустить собственную компьютерную программу и выяснить, согласуются ли результаты ее работы с опубликованным доказательством. Но, сколько бы раз мы это ни проделывали, нам все равно не удастся получить столь же убедительный результат, как, к примеру, приведенное мной доказательство того, что обрезанную шахматную доску невозможно полностью закрыть костяшками домино. Компьютерное доказательство невозможно воспринять целиком. Его не проверишь вручную, даже если проживешь миллиард лет. Хуже того, даже если бы это было возможно, никто не поверил бы результату. Человеку свойственно ошибаться, а за миллиард лет ошибок накопится множество.

Компьютеры, вообще говоря, не ошибаются. Если компьютер и человек параллельно проведут достаточно сложные арифметические вычисления и их результаты не сойдутся, то в подобном соревновании по-настоящему разумный человек всегда поставит на компьютер. Но в его работе нет определенности. Корректно функционирующий компьютер в принципе может совершить ошибку; к примеру, космическая частица, пролетев сквозь ячейку памяти, может изменить ее состояние с 0 на 1. От этого можно защититься, повторив расчет. Ошибиться может и разработчик компьютера, что гораздо серьезнее. Так, у процессора Intel P5 Pentium в стандартных операциях с плавающей точкой была ошибка: если его просили разделить 4 195 835 на 3 145 727, он выдавал в ответ 1,33373, тогда как верный ответ 1,33382. Как оказалось, четыре ячейки таблицы оставались незаполненными. Кроме того, причина компьютерной ошибки может крыться в операционной системе или в недостатках пользовательской программы.

Утверждение, что доказательство Аппеля–Хакена, полученное при помощи компьютера, изменило саму природу понятия «доказательство», вызвало горячие философские споры. Я понимаю, к чему клонят философы, но на самом деле концепция доказательства, которой пользуются профессиональные математики, отличается от той, что преподают студентам в курсе математической логики. Но даже если взять эту, более формальную концепцию, то ничто в ней не требует, чтобы логику каждого шага непременно проверял человек. Уже несколько столетий математики используют для рутинных арифметических операций механизмы. Даже если человек пройдется по доказательству с карандашом, проверяя каждую его строчку, и не обнаружит ошибок, то кто гарантирует нам, что он ничего не пропустил? Совершенная и безупречная логика — это идеал, к которому мы стремимся. Люди несовершенны; они делают, что могут, но полностью исключить элемент неопределенности невозможно.

Робин Уилсон в книге «Четырех красок достаточно» (Four Colours Suffice) точно сформулировал ключевой социологический аспект реакции общества:

> «Аудитория раскололась на два лагеря: тех, кому за 40, невозможно было убедить, что доказательство, проведенное компьютером, может быть верным, а тех, кому еще не исполнилось 40, невозможно было убедить, что верным может быть доказательство, содержащее 700 страниц вычислений вручную».

Если наши машины в чем-то превосходят нас, разумно их использовать. Могут измениться *методики* доказательства, но они и без компьютеров постоянно меняются: этот процесс и называется «исследованиями». При этом сама концепция доказательства не изменится радикально, если некоторые шаги вместо человека проделает компьютер. Доказательство — это рассказ; доказательство, полученное при помощи компьютера, — это рассказ, сюжет которого слишком длинен для подробного пересказа, и поэтому нам приходится довольствоваться кратким пересказом его основной линии и гигантским приложением в виде машинной распечатки.

После первой прорывной работы Аппеля и Хакена прошло уже немало времени, и математики привыкли к использованию компьютера. Они и сегодня *предпочитают* доказательства, основанные исключительно на человеческом разуме, но в большинстве своем уже не считают их единственно возможными. В 1990-е гг., правда, кое у кого еще были легкие оправданные сомнения в доказательстве Аппеля–Хакена, и некоторые математики решили повторить его целиком, воспользовавшись новыми теоретическими наработками и гораздо более мощными компьютерами. В 1994 г. Нил Робертсон, Дэниел Сандерс, Пол Сеймур и Робин Томас взяли из работы Аппеля–Хакена только базовую стратегию, отбросив все остальное, и повторили все с самого начала. За год им удалось найти неустранимый набор из 633 конфигураций, сводимость каждой из которых можно было доказать при помощи 32 правил разрядки. Результат оказался значительно проще, чем 1482 конфигурации и 487 правил разрядки Аппеля и Хакена. Сегодня компьютеры считают так быстро, что это доказательство можно целиком проверить на домашнем компьютере за несколько часов.

Все это, конечно, хорошо, но главным по-прежнему остается компьютер. Можно ли изменить ситуацию? В среде математиков зреет убеждение в том, что в данном случае это не исключено. Возможно, новые открытия, связанные с задачей о четырех красках, позволят когда-нибудь получить более простое доказательство. Для него не понадобится или почти не понадобится помощь компьютера, и математики смогут прочесть его, обдумать и сказать: «Да!» Пока такого доказательства нет, но что-то витает в воздухе…

Математики многое узнали о графах и сетях и узнают с каждым днем все больше. Топологи и геометры обнаруживают глубокие связи между сетями и совершенно далекими, казалось бы, от них областями математики, включая и некоторые разделы математической физики. Время от времени, скажем, всплывает концепция кривизны. Название ее говорит само за себя: кривизна пространства говорит о том, насколько это пространство изогнуто. Если оно плоское, как плоскость, его кривизна равна нулю. Если оно изогнуто в одну сторону — как холм во всех направлениях загибается вниз, — его кривизна положительна. Если пространство, как горный перевал, в некоторых направлениях загибается вниз, а в некоторых вверх, его кривизна отрицательна. Существуют геометрические теоремы (отдаленные потомки формулы Эйлера), связывающие построенные в пространстве сети с кривизной самого пространства. На это же намекает формула Хивуда для тора с g отверстиями. Сфера имеет положительную кривизну; тор, представленный в виде квадрата с тождественными противоположными сторонами (см. рис. 12 справа), имеет нулевую кривизну, а тор с двумя или более отверстиями — отрицательную. Так что между кривизной и раскрашиванием карт определенно существует какая-то связь.

За этой связью стоит одно полезное свойство кривизны: от нее очень сложно избавиться. Это похоже на кошку под ковром. Если ковер лежит на полу ровно, кошки под ним нет, но, если вы видите на ковре горб, значит, под ним кошка. Вы можете гонять эту кошку по всему ковру, но горб будет просто перемещаться с одного места на другое. Так же и кривизну можно сдвинуть, но невозможно убрать. Разве что кошка доберется до края ковра и выскочит наружу, унося кривизну с собой. Правила разрядки Хееша немного напоминают замаскированную кривизну. Они позволяют гонять электрический заряд с места на место, но не ликвидируют его. Не может ли существовать некое понятие кривизны для сети и хитрое правило разрядки, позволяющее, по существу, гонять по нему эту кривизну?

Если так, то нельзя исключить вариант, при котором вам удастся уговорить сеть раскраситься самостоятельно. Присвоить точкам (а может быть, и линиям тоже) кривизну, а затем позволить сети перераспределить ее более равномерно. Возможно, если мы все правильно подготовим, то «равномерность» будет означать как раз достаточность четырех красок. Это всего лишь идея, причем не моя, и я объяснил ее недостаточно подробно, чтобы что-нибудь понять. Но эта идея порождена интуицией какого-то математика и вселяет надежду на то, что в будущем, возможно, будет найдено более концептуальное доказательство теоремы о четырех красках — это будет потрясающая повесть, а не краткий пересказ с приложением в виде миллиарда телефонных справочников. В главе 10 мы столкнемся с аналогичной идеей в гораздо более хитроумном контексте, где она помогла решить еще более великую топологическую задачу.

5

Сферическая симметрия

Гипотеза Кеплера

Все началось со снежинки.

Снег обладает странной красотой. Он падает с небес пушистыми белыми хлопьями; он летит по ветру и образует мягкие холмы и гребни, покрывающие все вокруг; он сам по себе обретает неземные странные формы. Он холодный. По нему можно кататься на лыжах, на санках, можно лепить из него снежки и снеговиков… А если не повезет, можно оказаться погребенным под тысячетонной снежной лавиной. Исчезая, он не возвращается на небо — по крайней мере непосредственно в виде белых хлопьев. Он превращается в обычную воду, которая, конечно, может испариться и вернуться на небо, но может и течь ручьями и реками вниз, к морю, а потом долго-долго обитать в океане. Снег — форма льда, а лед — это замороженная вода.

Все сказанное не ново. Вероятно, это знали еще неандертальцы.

Снежинки ни в коем случае нельзя назвать бесформенными комками. В первозданном виде (до того, как начинается процесс таяния) многие из них представляют собой крохотные затейливые звездочки — плоские, шестигранные и симметричные. Есть и просто шестиугольники. Некоторые снежинки не настолько симметричны, некоторые отличаются заметной толщиной (т. е. имеют третье измерение), но снежинки с шестилучевой симметрией очень типичны и встречаются часто. Снежинки — кристаллы льда. Это тоже не новость, ведь невозможно, увидев кристалл, не узнать его. Но снежинки — не обычные кристаллы с плоскими гранями в виде многоугольников. Самое загадочное свойство снежинок добавляет в картину легкий штрих хаоса: несмотря на одинаковую симметрию, точная структура каждой снежинки уникальна. Говорят, не существует двух одинаковых снежинок. Я часто недоумевал: откуда они могут это знать? Но если достаточно педантично разобраться в том, что считать одинаковым, то выяснится, что цифры говорят в пользу такой позиции.

Почему для снежинок характерна шестилучевая симметрия? Этим вопросом 400 лет назад задался один из великих математиков и астрономов XVII в. — и после некоторых размышлений нашел на него ответ. Поразительно верный ответ, тем более что ученый при этом не проводил никаких особых экспериментов. Он просто свел воедино несколько простых общеизвестных мыслей. К примеру, о том, как располагаются зернышки в гранате.

Этого человека звали Иоганн Кеплер, и у него был очень хороший повод задуматься о снежинках. Жизнь и работа ученого в те времена часто зависели от богатого покровителя, и для Кеплера таковым был Иоганн Вакер фон Вакенфельс. В то время Кеплер служил придворным математиком императора Священной Римской империи Рудольфа II, а Вакер, дипломат, был советником императора. Кеплер хотел сделать своему покровителю новогодний подарок. В идеале он должен был быть недорогим, необычным и занимательным — и приоткрыть перед Вакером дверь в мир замечательных открытий, которые стали возможны благодаря его деньгам. Так что Кеплер собрал свои мысли о снежинках в небольшую книгу, которая и должна была стать подарком. Называлась она «О шестиугольных снежинках» (De Nive Sexangula). Шел 1611 г. В этой книге содержалось короткое замечание (один из основных шагов в рассуждениях Кеплера), и сформулированную в нем математическую загадку никому не удавалось решить на протяжении 387 лет.

Кеплер имел огромный опыт поиска закономерностей. Его важнейшая научная работа — открытие трех фундаментальных законов движения планет. Первый и самый известный из них гласит, что орбиты планет представляют собой эллипсы. Кроме того, Кеплер был мистиком и приверженцем учения пифагорейцев о том, что Вселенная основана на числах, закономерностях и математических формах. Помимо астрономии, он занимался астрологией: математики в те времена нередко выдавали себя за астрологов, поскольку обладали необходимой квалификацией и могли рассчитать, когда асцендент находится в Водолее. Богатые покровители и короли заказывали им составление гороскопов.

В своей книге Кеплер указал, что снег возникает из водяного пара, который сам по себе не имеет формы, но каким-то образом превращается в твердые снежинки с шестилучевой симметрией. Для такого перехода должна быть какая-то причина, считал Кеплер:

«Позволительно спросить, каково это действующее начало, как оно действует, является ли форма искони присущей телу или приобретается под влиянием внешних воздействий, принимает ли материал шестиугольную форму в силу необходимости или по своей природе и что следует считать врожденным: воплощенный в шестиугольном архетип красоты или знание цели, к которой приводит эта форма?»

В поисках ответа на этот вопрос Кеплер рассматривает другие примеры шестиугольных форм в природе. В первую очередь на ум приходят пчелиные соты. Они формируются из двух слоев шестиугольных ячеек, составленных «впритык», так что общие их концы образуют три ромба, т. е. параллелограммы, у которых все четыре стороны одинаковы. Такая форма напомнила Кеплеру о теле, известном как ромбический додекаэдр (см. рис. 15). Это тело не входит в перечень пяти правильных тел, известных пифагорейцам (их систематизировал Евклид), но обладает интересным отличительным свойством: одинаковыми ромбическими додекаэдрами можно полностью, без зазоров заполнить пространство. Эта же форма возникает внутри граната, где маленькие округлые зернышки растут в тесноте и потому вынуждены «изобретать» эффективную упаковку.

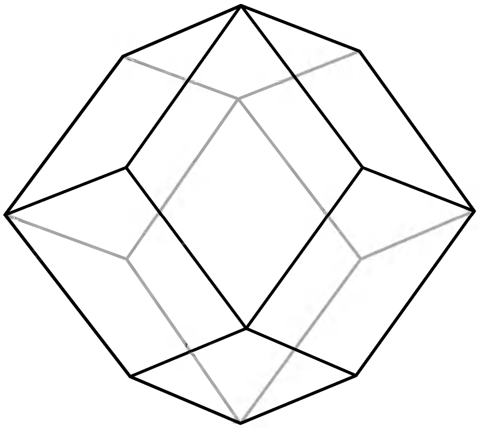


Рис. 15. Ромбический додекаэдр — тело с 12 ромбическими гранями

Как всякий разумный математик, Кеплер начинает с простейшего случая, при котором сферы (шарики) образуют единственный плоский слой. По существу, это то же самое, что упаковывать одинаковые кружочки на плоскости. Он находит всего две правильные конфигурации. В одной круги укладываются в квадраты (см. рис. 16 слева); в другой — в равносторонние треугольники (см. рис. 16 справа). Эти конфигурации, повторяемые на бесконечной плоскости, образуют квадратную решетку и треугольную решетку. Слово «решетка» говорит об их пространственной периодичности, повторяющейся в двух независимых направлениях. На рисунке по объективным причинам показаны лишь конечные участки решетки, поэтому на края не нужно обращать внимания. То же можно сказать про размещенные ниже рис. 17–20. На рис. 16 слева и справа показано по пять рядов кругов, и в каждом ряду они соприкасаются с соседями. Однако треугольная решетка немного сплюснута, и ее ряды располагаются ближе друг к другу. Так что круги в треугольной решетке упакованы плотнее, чем в квадратной.

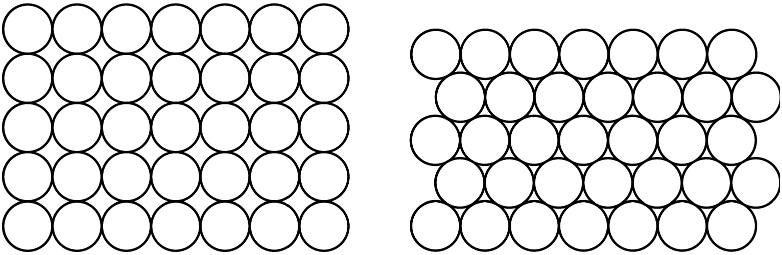


Рис. 16. Квадратная упаковка *(слева)*. Треугольная (иначе ее называют гексагональной, или шестиугольной) упаковка *(справа)*

Далее Кеплер задается вопросом, как такие слои можно уложить один на другой, и рассматривает четыре случая. В первых двух все слои уложены по квадратной решетке. При укладывании в стопку шарики каждого ряда можно поместить точно над шариками нижнего ряда. Тогда у каждого шарика будет по шесть непосредственных соседей: четыре в своем слое, один сверху и один снизу. Такая упаковка похожа на трехмерную шахматную доску, сделанную из кубиков; в нее она, кстати, и превратится, если надуть шарики так, чтобы дальше расширяться им было уже некуда. Но это, говорит Кеплер, «не самая плотная упаковка». Ее можно сделать еще плотнее, если сдвинуть верхний слой по диагонали так, чтобы его шарики точно легли во впадины между шариками нижнего слоя (см. рис. 17 слева). Повторим этот процесс для всех слоев (см. рис. 17 справа). Теперь у каждого шарика по 12 соседей: четыре в своем слое, четыре вверху и четыре внизу. Если их надуть, пространство заполнится ромбическими додекаэдрами.

В двух других случаях слои складываются по гексагональной решетке. Если при складывании в стопку поставить шарик над шариком, у каждого шарика будет по восемь соседей: шесть в своем слое, один вверху и один внизу. Опять же шарики верхнего слоя можно поставить над промежутками в нижнем. Тогда у каждого из них будет по 12 соседей: шесть в собственном слое, три вверху и три внизу. Количество соседей такое же, как во втором варианте упаковки квадратных слоев, и Кеплер, проведя тщательный геометрический анализ, показывает, что в реальности этот вариант упаковки полностью совпадает со вторым. Единственная разница заключается в том, что квадратные слои лежат не горизонтально, а под углом. Кеплер пишет: «Таким образом, самая плотная трехмерная упаковка с треугольной решеткой не может существовать без квадратной решетки, и наоборот». К этому я еще вернусь: это важно.

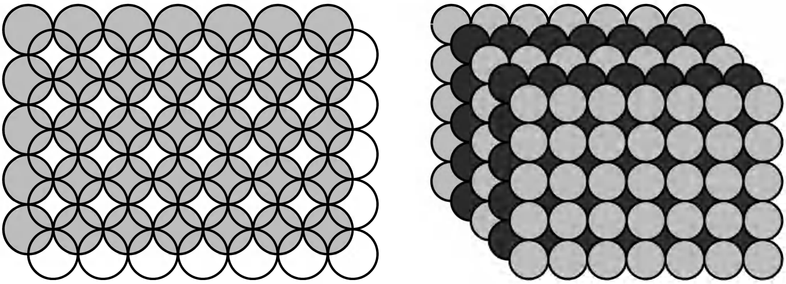


Рис. 17. Добавление второго слоя шариков (прозрачные кружки) поверх первого слоя (серые кружки) *(слева).* Повторение конструкции *(справа)*

Разобравшись с базовой геометрией упаковки шариков, Кеплер возвращается к снежинке с ее шестилучевой симметрией. Шестигранник напоминает ему о треугольной решетке упаковки шариков на плоскости, в которой каждый шарик соседствует с шестью другими, образуя идеальный шестиугольник. В этом, делает вывод Кеплер, должно быть, и заключается причина шестиконечности снежинок.

Эта глава посвящена, строго говоря, не снежинкам, но данное Кеплером объяснение их симметричности очень похоже на то, что предложили бы сегодня мы, так что стыдно было бы остановиться на этом. Почему они такие разные, но при этом все симметричны? Когда вода кристаллизуется, образуя лед, атомы водорода и кислорода, из которых состоят молекулы воды, укладываются в симметричную структуру — кристаллическую решетку. Эта решетка сложнее любой кеплеровой конструкции из шариков, но строится тоже на основе шестилучевой симметрии. Снежинка растет от крохотного «зернышка» всего из нескольких атомов, организованных в виде маленького кусочка решетки. Это зернышко тоже обладает шестилучевой симметрией; именно оно подготавливает сцену для роста ледяного кристалла в грозовой туче, где ветер бросает миллионы этих кристаллов из стороны в сторону.

Великое разнообразие узоров в снежинках — следствие меняющихся внешних условий. В зависимости от температуры и влажности рост кристалла может идти равномерно по всей границе, и тогда атомы с любой стороны добавляются с одной и той же частотой, и получаются простые шестиугольники. Но рост может идти с разной скоростью в разных местах, и тогда получается древовидная структура. Растущая снежинка путешествует по грозовой туче то вверх, то вниз, и условия вокруг постоянно меняются, причем случайным образом. Но сама снежинка настолько мала, что условия во всех шести ее углах в любой момент времени практически одинаковы. Поэтому все шесть лучей делают одно и то же. Каждая снежинка несет на себе отпечаток своей истории. На практике шестилучевая симметрия никогда не бывает строгой, но часто очень близка к идеалу. Лед — загадочное вещество, поэтому возможны и другие формы — пики, плоские круги, шестигранные призмы, призмы с плоскими концами. Полное описание происходящего вышло бы очень сложным, но определяющим фактором является то, как располагаются атомы в кристаллах льда. Во времена Кеплера атомная теория сводилась в лучшем случае к неопределенным предположениям древних греков. Поразительно, как далеко он сумел зайти в своих выводах, опираясь только на народные наблюдения, мысленные эксперименты и собственную интуицию.

Гипотеза Кеплера не имеет отношения к снежинкам как таковым. Все дело в небрежном замечании о том, что укладка слоев из плотно упакованных шариков, при которой шарики верхнего слоя ложатся во впадины между шариками нижнего, дает «самую плотную возможную упаковку в трех измерениях». Неформально эту гипотезу можно сформулировать так: если вы хотите упаковать много апельсинов в большой ящик, заполнив его при этом как можно плотнее, то укладывать плоды нужно так, как это делает любой торговец фруктами.

Трудность здесь не в том, чтобы найти *ответ*. Кеплер нам все рассказал. Трудность в том, чтобы доказать, что он был прав. За прошедшие столетия ученые собрали немало косвенных тому свидетельств. Никто не смог предложить более плотную упаковку. Именно такое расположение атомов часто встречается в кристаллах, где, как считается, плотность оптимальна для минимизации затрат энергии — это стандартный принцип, по которому созданы многие природные формы. Этого оказалось достаточно, чтобы убедить большинство физиков. И никто не смог доказать, что ничего лучшего *не существует*. В более простых вопросах такого рода, вроде упаковки кругов на плоскости, обнаружились скрытые глубины. Надо сказать, весь этот раздел математики сложен и полон неожиданностей. Все это тревожило математиков, хотя большинство из них тоже считали, что Кеплер дал верный ответ. В 1958 г. Амброз Роджерс описал гипотезу Кеплера как то, «во что многие математики верят, а все физики знают и так». В этой главе рассказывается, как математики обратили веру в точное знание.

Чтобы понять, что именно они сделали, нам придется как следует приглядеться к кеплеровой конструкции из шариков, известной как гранецентрированная кубическая решетка. Стоит сделать это, и начинают проявляться тонкости стоявшей перед математиками задачи. Первым на ум приходит вопрос: почему мы используем слои с квадратной решеткой? В конце концов, самой плотной упаковкой на плоскости (т. е. для одного слоя) является *треугольная* решетка. Дело в том, что гранецентрированную кубическую решетку можно получить и из слоев с треугольной укладкой шариков; именно в этом суть замечания Кеплера о том, что «треугольная схема укладки не может существовать без квадратной». Однако гранецентрированную кубическую решетку, сложенную из квадратных слоев, проще описывать. Кроме того, так мы убедимся, что гипотеза Кеплера не столь прямолинейна, как укладка апельсинов в ящики.

Предположим, что мы начинаем с плоского слоя шариков, уложенных треугольниками (см. рис. 16 справа). Между шариками имеются скругленные треугольные выемки, в которые могут лечь шарики следующего слоя. Когда мы начинали с квадратного слоя, мы могли использовать все выемки без исключения, и положение второго и последующих слоев определялось однозначно. С треугольными слоями не так. Мы не можем использовать все выемки, поскольку они располагаются слишком близко друг к другу. Мы можем использовать только половину. Один из вариантов укладки показан на рис. 18 слева при помощи небольших серых точек для наглядности, а рис. 18 справа демонстрирует, как следует расположить следующий слой шариков. Другой способ уложить новый слой в выемки предыдущего показан на рис. 19 слева темными точками. Эти точки совпадают с выемками второго слоя, так что мы добавляем третий слой в соответствующем положении. Результат показан на рис. 19 справа.

Если мы работаем всего лишь с двумя слоями, разница между двумя вариантами не играет никакой роли. Мы можем без труда получить первый вариант укладки, просто повернув второй вариант на 60°. Эти варианты одинаковы «с точностью до симметрии». Но после укладки первых двух слоев у нас появляются два по-настоящему разных варианта для третьего слоя. Каждый новый слой имеет две системы выемок, показанных на рис. 19 слева светлыми и темными точками. В одной из них выемки соответствуют центрам шариков предыдущего слоя, которые на рис. 19 справа видны как светло-серые треугольнички. Во второй выемки соответствуют выемкам предпредыдущего слоя и видны на рис. 19 справа как треугольнички с вписанными в них крохотными белыми шестиугольничками. Чтобы получить гранецентрированную кубическую решетку, мы должны использовать для третьего слоя темно-серые позиции, а затем повторять такой порядок укладки до бесконечности.

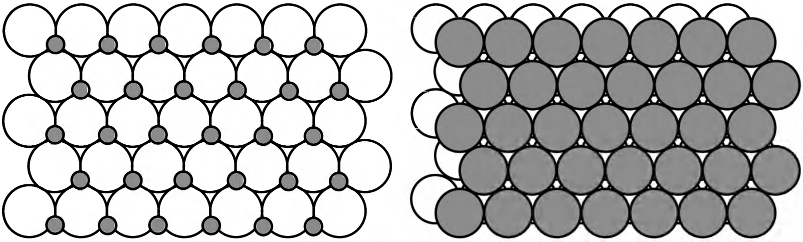


Рис. 18. Укладка треугольной решетки в систему выемок предыдущего слоя

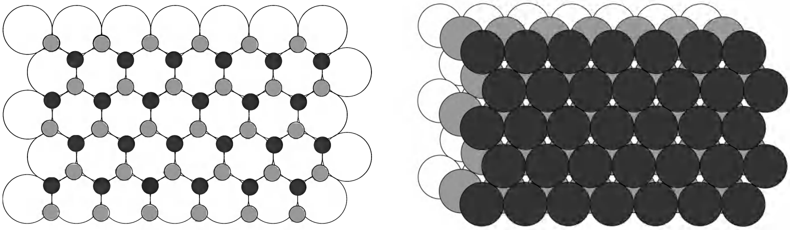


Рис. 19. Укладка треугольных слоев один на другой

Не до конца очевидно, однако, что результатом такой укладки станет гранецентрированная кубическая решетка. Где же здесь квадраты? Дело в том, что квадраты в такой укладке присутствуют, но располагаются наклонно, под углом. На рис. 20 показаны шесть последовательных треугольных слоев, из которых удалена часть шариков. Стрелками указаны ряды и столбцы скрытой внутри квадратной решетки. Все слои, параллельные данному, тоже выстроены по квадратной решетке, а между собой соотносятся в точности так же, как я выстраивал гранецентрированную кубическую решетку.

Насколько компактна такая упаковка? Мы измеряем компактность (эффективность) упаковки ее плотностью: долей общего объема, занимаемой шариками[20]. Чем больше плотность, тем компактнее упаковка. Кубики укладываются в параллелепипед с плотностью 1, заполняя весь объем. Между шариками, очевидно, в любом случае останутся промежутки, так что плотность их упаковки меньше единицы. Плотность гранецентрированной кубической решетки составляет в точности $^{\pi}/_{18}$, это примерно 0,7405. При такой упаковке шарики заполняют чуть меньше трех четвертей пространства, и гипотеза Кеплера утверждает, что никакая упаковка шариков не может иметь плотность больше этой.

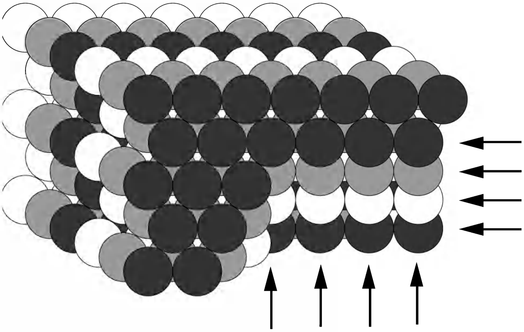


Рис. 20. Внутри стопки треугольных слоев скрываются наклонные квадратные слои

Я сформулировал все это достаточно осторожно. Я не сказал, что «плотность гранецентрированной кубической решетки выше, чем любой другой». Такое утверждение было бы неверным, и в этом несложно убедиться. Для этого вернемся к построению гранецентрированной кубической решетки из треугольных слоев. Я сказал, что после укладки первых двух слоев возникает два варианта укладки третьего. Гранецентрированная кубическая решетка возникает во втором варианте — том, что с темно-серыми точками. Что произойдет, если мы пойдем по первому пути и используем светло-серые точки? Тогда шарики третьего слоя окажутся точно над шариками первого. Продолжив точно так же и помещая каждый новый слой точно над позапрошлым слоем, мы получим второй вариант объемной решетки: гексагональную. Она отличается от гранецентрированной кубической решетки, но имеет ту же плотность. Интуитивно это понятно, поскольку два разных способа укладки третьего слоя симметричны относительно поворота, а сам слой в обоих случаях ложится на предыдущий одинаково плотно.

Это единственные два способа *решетчатой* упаковки, которые можно получить при укладке стопки треугольных слоев, но в 1883 г. географ и кристаллограф Уильям Барлоу заметил, что для каждого следующего слоя мы можем произвольно выбрать любой из двух вариантов укладки. Поскольку оба варианта вносят в плотность всей стопки одинаковый вклад, плотность всех этих вариантов упаковки будет одинакова и равна $\pi/\quad 18$. При этом существуют бесконечно много случайных последовательностей такого рода и, соответственно, бесконечно много различных вариантов упаковки с одинаковой плотностью.

Короче говоря, нет единственной самой плотной объемной упаковки шариков. Их бесконечно много, и все они одинаково плотные. Отсутствие единственно верного решения — предупреждение: проблема не так проста и прямолинейна. Если Кеплер был прав, то существует оптимальная *плотность* упаковки, но есть и бесконечное множество различных структур, ею обладающих. И чтобы доказать оптимальность этой плотности, недостаточно успешно пристраивать каждый новый шарик к предыдущим как можно плотнее. Есть варианты.

Конечно, торговцы фруктами обладают невероятно богатым опытом — ведь гранецентрированную кубическую решетку наверняка можно было увидеть на рынках Древнего Египта еще в додинастическую эпоху, — но одним лишь опытом в таком деле никак не обойдешься. Вообще, тот факт, что метод торговцев фруктами дает хороший результат, в определенной мере случайность. Задача торговца фруктами состоит не в том, чтобы упаковать апельсины как можно плотнее в пространстве, где возможна, в принципе, любая конструкция. Его задача — уложить плоды как можно надежнее в мире, где земля плоская, а сила тяжести действует сверху вниз. Поэтому торговец начинает с того, что выкладывает апельсины в один слой — это очень естественно; затем он добавляет сверху еще один слой и т. д. Если ящик, в который укладываются плоды, прямоугольный, то первый слой, скорее всего, будет выложен по квадратной решетке. Если площадь ничем не ограничена, то естественной будет либо квадратная, либо треугольная решетка. В конечном итоге обе дают гранецентрированную кубическую решетку — по крайней мере, если треугольные слои укладываются как следует. Вообще говоря, квадратная решетка представляется не лучшим вариантом — ведь это не самый плотный способ укладки одного слоя. Однако — скорее по счастливой случайности, чем в результате осознанного выбора — это, как оказалось, не имеет значения.

Физики не интересуются апельсинами, их больше занимает то, как соседствуют друг с другом атомы. Кристалл — это регулярная, пространственно периодическая конструкция из атомов. Гипотеза Кеплера утверждает, что периодичность кристалла — это естественное следствие максимально плотной «упаковки» атомов. Для большинства физиков само существование кристаллов является достаточным доказательством, — по их мнению, гипотеза Кеплера очевидно верна. Однако мы только что убедились, что существует бесконечно много способов упаковки шариков с точно такой же плотностью, как у гранецентрированной кубической и гексагональной решеток, и что способы эти не являются пространственно периодическими. Так почему в кристаллах природа использует именно периодические структуры? Возможно, ответ в том, что атомы не следует рассматривать как сферические объекты.

Математики тоже не слишком интересуются апельсинами. Подобно Кеплеру, они предпочитают работать с идеальными и идентичными сферами, и доводы физиков не представляются им убедительными. Судите сами: если при моделировании кристаллов атомы не следует рассматривать как идеально круглые шарики, то существование кристаллов вообще не имеет отношения к гипотезе Кеплера и ничего не доказывает. Либо то, либо другое. Даже если вы скажете, что гипотеза как будто объясняет кристаллическую решетку, а кристаллическая решетка как будто показывает, что гипотеза верна… все равно в таком рассуждении будет логический пробел. Математикам нужно доказательство.

Кеплер не называл свое утверждение гипотезой: он просто высказал его в своей книге. Совершенно неясно, собирался ли он интерпретировать упомянутый факт столь всеобъемлющим образом. Имел ли он в виду, что гранецентрированная кубическая решетка представляет собой «самую плотную упаковку в трех измерениях» из всех представимых способов упаковки шариков? Или просто говорил о том, что это самая плотная упаковка из рассмотренных им лично? Невозможно вернуться в прошлое и спросить об этом. Но, как бы ни обстояло тогда дело, математиков и физиков интересует именно общая, самая смелая формулировка. Та, что требует рассмотреть все возможные способы упаковки бесконечного числа шариков в бесконечном пространстве и показать, что ни один из этих способов не может похвастать большей плотностью, чем гранецентрированная кубическая решетка.

Недооценить сложность гипотезы Кеплера очень легко. Вроде бы логично предположить, что самая плотная упаковка получится, если добавлять шарики по одному, так, чтобы каждый из них касался как можно большего числа соседних. Такой подход непременно даст структуру, о которой говорил Кеплер. То же получится, если вы будете добавлять шарики в правильном порядке и всегда, когда есть альтернативы, выбирать для них верную позицию. Однако нет никакой гарантии, что более дальновидная политика не окажется лучше, чем процесс поштучного добавления шариков. Всякий, кому приходилось укладывать вещи в багажник автомобиля, знает, что при укладке их по одной в багажнике могут остаться промежутки, куда ничего больше не лезет, но если начать сначала и подойти к вопросу более тщательно, то иногда удается втиснуть в то же пространство больше вещей. Конечно, отчасти проблема укладки вещей затрудняется тем, что все они имеют разные размеры и форму, но смысл аналогии достаточно понятен: максимально плотная упаковка на одном небольшом участке пространства может затруднить укладку остальных вещей и не привести к максимально плотной упаковке в большем объеме.

Конструкции, которые рассматривает Кеплер, очень специфичны. Можно предположить, что какой-то совершенно иной принцип позволит упаковать одинаковые шарики еще плотнее. Может быть, выпуклые слои были бы более эффективны. А может быть, «слои» — вообще неудачная идея. Но даже если вы абсолютно убеждены, что все сделано правильно, это все равно нужно доказывать.

Не убеждены? По-прежнему считаете, что здесь все очевидно? Настолько очевидно, что никакого доказательства не требуется? Сейчас я попытаюсь разрушить вашу уверенность в правильности интуитивного решения — на более простом примере, где речь идет об укладке одинаковых кружочков на плоскости. Предположим, я дам вам 49 одинаковых кружочков единичного диаметра. Каким будет размер *самого маленького* квадрата, способного их все вместить без перекрытия? На рис. 21 слева показан очевидный ответ: расположить их, как ставят молочные бутылки в ящике. Сторона квадрата при этом — ровно 7 единиц. Чтобы убедиться, что это наилучший вариант, обратите внимание на то, что каждый кружок жестко удерживается остальными, так что лишнее место взять неоткуда. Но рис. 21 справа показывает, что этот ответ неверен. Стоит упаковать кружочки вот таким немного нерегулярным образом, и они поместятся в квадрате со стороной чуть меньше 6,98. Так что доказательство тоже неверно. Жесткость упаковки не гарантирует, что невозможно сделать плотнее.

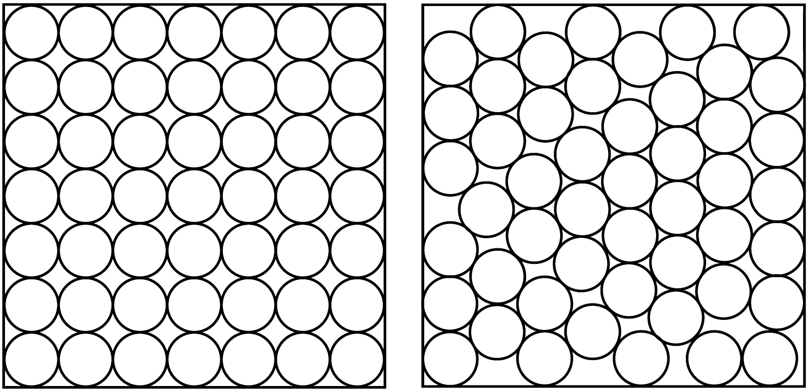


Рис. 21. 49 кружков в квадрате 7 × 7 (слева). А так можно разместить 49 кружков в квадрате чуть меньшего размера (справа)

Несложно убедиться, что рассуждения, позволяющие получить ответ «семь», просто не могут быть верными. Для этого достаточно рассмотреть квадрат побольше. Квадратная решетка позволяет поместить n^2 кружков единичного диаметра в квадрат со стороной n. Невозможно повысить плотность такой укладки путем плавного перемещения кругов, ведь укладка-то у нас жесткая. Но для больших n должны существовать и более плотные укладки, потому что, как известно, треугольная решетка эффективнее квадратной. Если взять по-настоящему большой квадрат и упаковать в него как можно больше кругов, используя треугольную решетку, то, в конце концов, треугольная решетка, благодаря своим преимуществам, победит, несмотря на «краевые эффекты» по границе квадрата, где придется оставлять незаполненные промежутки. Периметр квадрата — $4n$ — невелик по сравнению с n^2. Треугольная решетка одерживает верх над квадратной как раз при $n = 7$. Это не очевидно, и доказывать это пришлось бы долго и подробно, но ясно, что рано или поздно размер сработает. Одной только жесткости укладки недостаточно.

На самом деле существует два варианта гипотезы Кеплера. Одна рассматривает только регулярную упаковку; в ней центры шариков образуют пространственную периодическую структуру, которая бесконечно повторяется в трех независимых направлениях, как своего рода объемные обои. Даже в этом случае проблема весьма сложна, поскольку в пространстве возможно множество различных решеток. Кристаллографы различают 14 их типов, различаемых по видам симметрии, и некоторые из этих типов имеют параметры, которые могут принимать бесконечно много различных значений. Но все эти сложности меркнут, когда начинается рассмотрение второго варианта гипотезы, разрешающей любые возможные упаковки. Шарики здесь находятся в пространстве без гравитации и совершенно не обязаны собираться в слои или какие бы то ни было симметричные структуры.

Когда задача представляется слишком сложной, математики обыкновенно убирают ее в долгий ящик и принимаются искать более простые варианты. Рассуждения Кеплера о плоских слоях наводят на мысль о более простой задаче — упаковке кругов на плоскости. Это значит, что на плоскости, где имеется бесконечное число одинаковых кругов, надо собрать их как можно плотнее. В такой задаче плотность — это доля *площади,* которую покрывают круги. В 1773 г. Жозеф Луи Лагранж доказал, что самая плотная упаковка кругов на плоскости достигается в треугольной решетке, где плотность составляет $\pi/\quad 12 \approx$ $\approx 0{,}9069$. В 1831 г. Гаусс в отзыве на книгу Людвига Зибера (тот обобщил некоторые теоретические выводы Гаусса по уравнениям третьего порядка) отметил: результаты Зибера доказывают, что самую плотную регулярную упаковку в трехмерном пространстве обеспечивают гранецентрированная кубическая и гексагональная решетки. Сегодня математики очень много знают о решетчатой укладке в пространствах с бóльшим числом размерностей — четыре, пять, шесть и т. д. Особенно хорошо удалось разобраться с решетками в 24-мерном пространстве. (Да, такая уж это тема.) Несмотря на кажущуюся непрактичность, эта область математики имеет немало применений в теории информации и теории кодирования.

Нерешетчатые укладки — совершенно иное дело. Их бесконечно много, и они не отличаются приятной регулярностью структуры. Почему бы нам не удариться в другую крайность и не попробовать случайную укладку? Стивен Гейлс в своем труде «Статика растений» (1727) рассказал об экспериментах, в ходе которых он «вжимал несколько пакетов свежего гороха в один горшок» и обнаруживал, что при сильном сдавливании они образуют то ли «красивые правильные додекаэдры», то ли «достаточно правильные додекаэдры»[1]. Судя по всему, автор говорил о том, что правильные додекаэдры красивы, а не о том, что додекаэдры получались довольно правильные, но вторая интерпретация лучше, потому что настоящими правильными додекаэдрами невозможно без пустот заполнить пространство. Вероятно, он видел перед собой ромбические додекаэдры, которые, как мы уже убедились, связаны с гранецентрированной кубической решеткой. Дэвид Скотт насыпа́л в контейнер множество шариков от подшипников, тщательно встряхивал и измерял плотность укладки. По его данным, максимально она равнялась 0,6366. В 2008 г. Сун Чаомин, Ван Бин и Эрнан Макс получили эту же величину аналитически. Однако их результат не подразумевает, что Кеплер был прав, хотя бы потому, что это означало бы, что гранецентрированная кубическая решетка с плотностью 0,74 не может существовать. Простейшее объяснение такого расхождения заключается в том, что их результат не учитывает чрезвычайно редкие исключения, при том что и гранецентрированная кубическая, и гексагональная решетки, и все бесчисленные варианты конструкций из треугольных слоев представляют собой имен-

---

[1] Выражение pretty regular dodecahedrons в переводе с английского может означать как то, так и другое. — *Прим. пер.*

но такие исключения. По тому же принципу может существовать и еще какая-нибудь конструкция с еще большей плотностью. Это не может быть регулярная конструкция, но найти ее при помощи бессистемного поиска невозможно, ибо ее вероятность равна нулю. Так что исследование случайных вариантов упаковки, хотя и важно для многих областей физики, не слишком много говорит нам о гипотезе Кеплера.

Первый по-настоящему важный прорыв в решении этой задачи произошел в 1892 г., когда Аксель Туэ в ходе лекции на Скандинавском конгрессе естественных наук коротко изложил доказательство того, что никакая упаковка кругов на плоскости не может быть плотнее треугольной решетки. Лекция была опубликована, но формулировки там слишком неопределенны, чтобы можно было реконструировать само доказательство. Новую версию того же доказательства Туэ опубликовал в 1910 г. Оно казалось убедительным, за исключением нескольких технических моментов, которые, как считал автор, можно было без особого труда довести до ума. Но Ласло Фейеш Тот, вместо того чтобы заполнять пробелы в чужом доказательстве, получил в 1940 г. собственное, полное, основанное на других методах. Вскоре после этого Беньямино Сегре и Курт Малер представили альтернативные доказательства. А в 2010 г. Чан Хайчау и Ван Личун выложили в Интернет более простое доказательство.

Задача поиска наибольшей плотности укладки кругов или шариков при определенных условиях относится к классу математических задач, известных как задачи оптимизации. В такой задаче предлагается найти максимальное или минимальное значение некоторой функции (т. е. математического правила вычисления некой величины, которая определенным образом зависит от некоего набора переменных). Правило вычисления функции часто задается формулой, но это не обязательно. К примеру, таким образом можно сформулировать задачу с 49 кругами на плоскости. Переменными здесь будут координаты центров всех 49 кругов, а поскольку на каждый круг потребуется по две координаты, всего переменных получится 98. Сама функция — это размер наименьшего квадрата со сторонами, параллельными координатным осям, в который можно поместить данный набор неперекрывающихся кругов. Задача эквивалентна поиску минимального значения, которое принимает эта функция при значениях переменных, соответствующих всем вариантам решетки.

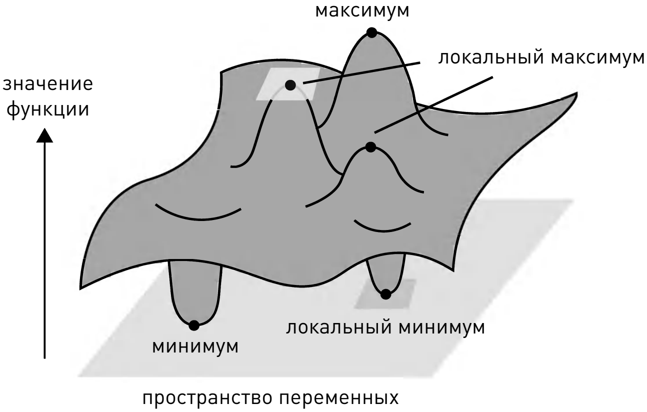


Рис. 22. Горы и долины функции

Функцию можно представить в виде многомерного ландшафта, каждая точка на котором соответствует определенному набору переменных, а высота в этой точке — значению функции. Максимум функции — это высота самого высокого ее пика, а минимум — глубина самой глубокой долины. В принципе задачи оптимизации можно решать методами дифференциального исчисления: в максимуме или минимуме функция должна быть горизонтальна (см. рис. 22), и дифференциальное исчисление позволяет отразить это условие в уравнении. Чтобы решить задачу укладки кругов в квадрат этим методом, нам пришлось бы решить систему из 98 уравнений с 98 переменными.

Решение задач оптимизации встречает на пути одно неожиданное препятствие: подобные уравнения часто имеют большое количество решений. Ландшафт может иметь множество локальных максимумов, из которых самый высокий лишь один. Представьте себе, к примеру, Гималаи: кроме пиков, там почти ничего и нет, но лишь Эвересту принадлежит рекорд высоты. Методы поиска пиков, самый очевидный из которых звучит как «иди вверх, пока это возможно», часто выводят на локальные максимумы и застревают на них. Еще одна трудность состоит в том, что с ростом числа переменных растет и вероятное число локальных пиков. Тем не менее иногда такой метод срабатывает. Даже частичные результаты могут оказаться полезными: если вам удалось найти локальный пик, ясно, что настоящий максимум не может быть ниже. Именно так была найдена улучшенная раскладка кругов в квадрате.

Для регулярных укладок функция, максимум которой нужно найти, зависит от конечного числа переменных — направлений и длин, вдоль которых решетка повторяется. Для нерегулярных укладок функция зависит от бесконечно большого числа переменных — центров всех кругов или шариков. В подобных случаях прямое использование дифференциального исчисления и других методик оптимизации ничего не даст. Доказательство Тота основано на хитрой идее переформулировать задачу о нерегулярной упаковке кругов и превратить ее в задачу оптимизации с *конечным* набором переменных. Позже, в 1953 г., он понял, что тот же трюк в принципе можно проделать и с гипотезой Кеплера. К несчастью, получившаяся функция зависит примерно от полутора сотен переменных — слишком много для ручного расчета. Но Тот прозорливо разглядел возможный выход: «Имея в виду стремительное развитие наших компьютеров, можно предположить, что минимум можно будет определить с высокой точностью».

В то время вычислительная техника только начинала развиваться, и достаточно мощной машины попросту не существовало, так что в последующие годы работа над гипотезой Кеплера шла в других направлениях. Ряд математиков занимался уточнением верхней границы для возможного значения плотности сферической упаковки. К примеру, в 1958 г. Роджерс доказал, что плотность не превосходит 0,7797. И никаких исключений: эта оценка относилась к любым способам укладки шариков. В 1986 г. Дж. Линдси понизил этот предел до 0,77844, а Дуглас Мадер в 1988 г. чуть-чуть улучшил оценку и получил 0,77836. Эти результаты показали, что невозможно получить плотность *намного* выше, чем 0,7405, характерные для гранецентрированной кубической решетки. Но тем не менее пробел в доказательстве сохранялся.

В 1990 г. американский математик Ву-И Хзянь объявил, что гипотеза Кеплера доказана. Подробности были опубликованы, но сразу появились сомнения. Тот, просмотрев статью в журнале *Mathematical Reviews,* написал: «Если бы спросили меня, [доказана ли тем самым гипотеза Кеплера] я бы ответил: нет. Надеюсь, что Хзянь дополнит свое сообщение, но мне кажется, что значительная часть работы еще впереди».

Томас Хейлс, посвятивший работе над гипотезой много лет, тоже высказал сомнения в том, что метод Хзяня можно довести до ума. Вместо этого он решил, что настало время всерьез отнестись к подходу, предложенному Тотом. Выросло новое поколение математиков, для которых естественнее обратиться к компьютеру, чем к таблице логарифмов. В 1996 г. Хейлс обрисовал стратегию доказательства, основанную на идее Тота. Эта стратегия требовала определить все возможные способы организации шариков в непосредственной близости от данного шарика. Укладка шариков определяется центрами соответствующих сфер. Для единичных сфер расстояние между центрами должно составлять по крайней мере две единицы. Скажем, что два шарика являются *соседями,* если их центры разделены расстоянием не более 2,51 единицы. Этот предел следует установить аккуратно: при слишком маленьком расстоянии не хватит места для перестановки шариков и повышения плотности, а при слишком большом количество вариантов расстановки соседних шариков станет просто гигантским. Хейлс выяснил, что 2,51 — это эффективный компромисс. Теперь мы можем представить, как располагаются шарики-соседи, и построить в пространстве бесконечную сеть. Узлами в ней станут центры шариков, а соединяться между собой линиями они будут, если соответствующие шарики — соседи. Эта сеть — своеобразный скелет укладки — содержит всю существенную информацию о соседях каждого шарика.

Для любого шарика мы можем взять непосредственно соседствующие с ним шарики и рассматривать далее только линии между ними, исключив сам первоначальный шарик. В результате получим своего рода клетку, или ячейку, вокруг точки в центре первоначальной сферы. На рис. 23 показаны: (левая пара) соседи шарика в гранецентрированной кубической решетке и соответствующая ячейка; (правая пара) то же самое для особой структуры из шариков — пятиугольной призмы; эта структура оказалась ключевым игроком всего доказательства. Вы видите здесь два плоских пятиугольника, параллельные «экватору» центрального шарика и еще по шарику на каждом полюсе конструкции.

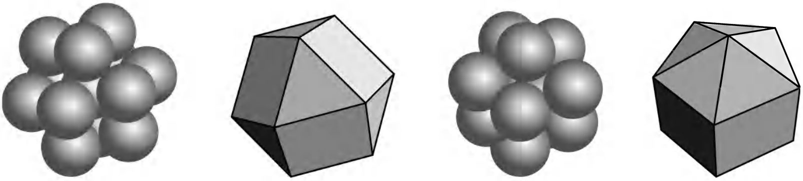


Рис. 23. *Слева направо:* ближайшие окрестности шарика в гранецентрированной кубической решетке; ячейка, образованная соседями такого шарика; окрестности шарика в решетке с пятиугольными призмами; ячейка, образованная соседями такого шарика

Ячейку можно изобразить как твердое тело с плоскими гранями; при этом плотность укладки вблизи центрального шарика будет определяться геометрией этого тела[21]. Ключевая идея состоит в том, чтобы каждой ячейке поставить в соответствие некое число (назовем его мерой ячейки), которое станет способом оценки плотности упаковки соседей центрального шарика. Мера ячейки — это не собственно плотность; это величина, которая ведет себя лучше и вычисляется проще плотности. В частности, меру ячейки можно найти просто как сумму мер ее граней; с плотностью такой метод не работает. В принципе, множество разных определений меры соответствуют этому условию, но все они сходятся в одном: для гранецентрированной кубической и гексагональной решеток мера, как бы вы ее ни определяли, всегда составляет восемь пунктов. В данном случае пункт равняется вполне конкретному числу:

$$4\text{arctg}\,\frac{2}{5}-\frac{\pi}{3}=0{,}0553736\cdot$$

Так что восемь пунктов на самом деле равняются 0,4429888. Это занятное число вычисляется исходя из особой геометрии гранецентрированной кубической решетки. Ключевое наблюдение Хейлса связывает гипотезу Кеплера с этим числом: если любая ячейка имеет меру восемь или меньше, гипотеза Кеплера верна. Таким образом, фокус доказательства смещается на ячейки и их меры.

Ячейки можно классифицировать по топологии: сколько у них имеется граней с тем или иным числом сторон и как эти грани сочетаются друг с другом. Однако при любой заданной топологии стороны граней ячейки могут быть различной длины. Длина сторон влияет на меру, но топология соединяет множество разных ячеек, и их можно рассматривать в общем. В своем доказательстве Хейлс рассмотрел около 5000 типов ячеек, но основные расчеты велись по нескольким сотням вариантов. В 1992 г. он предложил проект доказательства из пяти этапов:

1. Доказать утверждение в случае, когда все грани ячейки представляют собой треугольники.

2. Показать, что гранецентрированная кубическая и гексагональная укладки имеют более высокую меру, чем любая ячейка с той же топологией.
3. Разобраться со случаем, когда все грани ячейки представляют собой треугольники и четырехугольники, за исключением пятиугольной призмы (это более сложный случай).
4. Разобраться с ячейками, грани которых имеют более четырех сторон.
5. Разобраться с единственным оставшимся случаем — с ячейкой в виде пятиугольной призмы.

Часть 1 была реализована в 1994 г., часть 2 — в 1995 г. По ходу программы Хейлс модифицировал определение ячейки так, чтобы упростить рассуждения (он использовал термин «звезда декомпозиции»). Новое определение не изменило две представленные ячейки и почти не затронуло те части доказательства, которые были уже реализованы. К 1998 г. при помощи новой концепции были завершены все пять предложенных Хейлсом этапов. Часть 5 — хитрый случай с пятиугольной призмой — рассмотрел студент Хейлса Сэмюел Фергюсон.

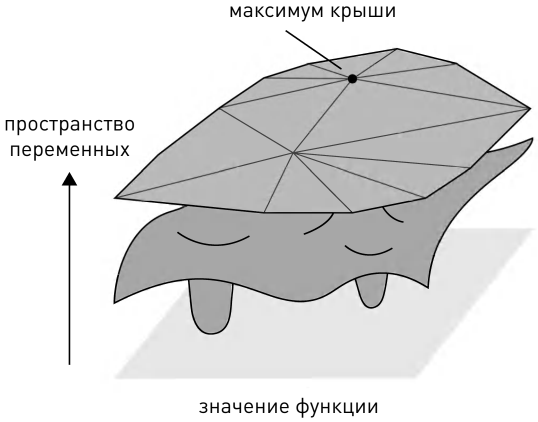


Рис. 24. Установка крыши над верхушкой функции

На всех этапах анализа были масштабно задействованы компьютеры. Хитрость здесь в том, чтобы для каждой локальной сети выбрать такое представление о мере, которое сделало бы вычисления сравнительно несложными. Геометрически замена плотности на меру выглядит как сооружение своеобразной крыши над гладким ландшафтом, максимум которого ищет исследователь. Крыша состоит из нескольких плоских участков (см. рис. 24). Работать с такими формами значительно легче, чем с гладкими поверхностями, поскольку максимумы располагаются в углах, для нахождения которых достаточно решить куда более простые уравнения. Для этого существуют эффективные методы, известные как линейное программирование. Если крыша построена так, что ее пик совпадает с пиком гладкой поверхности, то более простые расчеты, предназначенные для поисков пика крыши, позволят найти не только его, но и пик гладкой поверхности.

У всякого упрощения есть своя цена, и у этого тоже: чтобы найти пик крыши, приходится решать около 100 000 линейных задач. Впрочем, достаточно длинные расчеты, необходимые для этого, вполне по силам современным компьютерам. Когда Хейлс и Фергюсон подготовили свою работу к публикации, в ней оказалось около 250 страниц математического текста и 3 Гбайт компьютерных файлов.

В 1999 г. Хейлс подал подготовленное доказательство в *Annals of Mathematics,* и журнал специально для этого случая набрал жюри из 12 экспертов. К 2003 г. оно объявило, что «с уверенностью на 99%» представленное доказательство верно. Оставшийся 1% неуверенности касался компьютерных расчетов; многие из них жюри повторило, а также иными способами проверило стратегию и тактику доказательства, но некоторые его аспекты проверить не удалось. После значительной задержки журнал опубликовал работу. Хейлс признал, что такой подход к доказательству, вероятно, никогда не будет признан на 100% корректным, и в 2003 г. объявил о старте проекта по переформатированию доказательства в такой вид, в котором его сможет проверить компьютер при помощи стандартных программ автоматической проверки доказательств.

Это может показаться чем-то вроде перехода из огня в полымя, но на самом деле все предельно понятно и логично. Доказательства, которые математики публикуют в журналах, призваны убеждать людей. Как я уже говорил в главе 1, доказательство — это своего рода рассказ. Компьютеры не сильны в литературе, зато прекрасно справляются с заданиями, которые нам не по зубам: они способны безошибочно выполнять длинные нудные расчеты. Компьютеры идеально сочетаются с формальным определением доказательства в университетских учебниках: серия логических шагов, каждый из которых вытекает из предыдущих.

Компьютерщики научились использовать эту способность. Чтобы проверить доказательство, заставьте компьютер проанализировать каждый логический шаг. Звучит просто, но на самом деле доказательства в журналах пишут не так. Эти рассказы оставляют за скобками все рутинное или очевидное… Все привыкли к традиционным фразам: «Несложно убедиться, что…», «Используя методы Чизбургера и Чипса, модифицированные так, чтобы учитывать изолированные сингулярности, видим, что…», «Несложный расчет показывает…». Компьютеры (пока) с подобными задачами не справляются. Но люди-то всегда могут переписать доказательство, заполнив все подобные пропуски, и тогда компьютер вполне может проверить каждый его шаг.

Мы не прыгаем сразу же обратно в огонь по одной простой причине: программы, которые проверяют доказательство, тоже необходимо проверить, но лишь *один раз*. Вообще-то это универсальное программное обеспечение, которое применимо к любому доказательству, записанному в надлежащем формате. Все, что может вызывать сомнения, сконцентрировано здесь. Проверьте его — и потом с его помощью можно проверять все остальное. Можно даже упростить себе работу, написав программу проверки доказательства на языке, который позволит проверить ее при помощи гораздо более простой программы проверки.

В последние годы таким образом были проверены доказательства многих ключевых математических теорем. Для этого их нередко требовалось перевести в другую форму, более подходящую для компьютерных манипуляций. Один из последних триумфов — проверка и подтверждение доказательства теоремы Жордана: всякая замкнутая кривая без самопересечений на плоскости делит плоскость на две связные области. Утверждение может показаться очевидным, но пионеры топологии долго не могли строго доказать его. В конце концов это удалось в 1887 г. Камилю Жордану, опубликовавшему доказательство на 80 страницах, но позже его нередко критиковали за необоснованные ограничения. Поэтому слава досталась Освальду Веблену, давшему в 1905 г. более подробное доказательство. Веблен заявил: « [Жорданово] доказательство… не удовлетворяет многих математиков. В нем теорема принимается без доказательства в существенном особом случае, когда речь идет о простом многоугольнике; что же касается дальнейшего изложения, то следует признать по крайней мере, что не все детали в нем приведены».

Математики без колебаний приняли критику Веблена, но недавно Хейлс еще раз проанализировал доказательство Жордана и не нашел в нем «ничего, на что можно было бы возразить». Более того, замечание Веблена о многоугольнике звучит странно: теорема для него достаточно прозрачна, да и доказательство Жордана вовсе не опирается на этот частный случай[22]. У доказательств-рассказов есть собственные проблемы. С ними надо держать ухо востро и проверять, совпадает ли популярная версия рассказа с его оригинальным вариантом.

В процессе работы над гипотезой Кеплера Хейлс получил в 2007 г. формальное, проверенное компьютером доказательство теоремы Жордана, на что потребовалось 60 000 строк компьютерного кода. Вскоре после этого группа математиков, воспольproperty зовавшись другим программным обеспечением, получила другое формальное доказательство. Компьютерная проверка не застрахована от ошибок на 100%, но то же можно сказать и о традиционных доказательствах. Более того, многие математические научные труды, вероятно, содержат технические ошибки. Время от времени такие ошибки обнаруживаются и в большинстве случаев оказываются безвредными. Серьезные ошибки, как правило, замечают раньше, чем они приводят к нарушениям и делают что-то явно бессмысленным. Это еще один недостаток доказательства-рассказа — плата за то, что доказательство делается понятным человеку: иногда нестрогая логика выглядит на первый взгляд очень убедительно.

Хейлс называет свой подход Project FlysPecK. Согласно первоначальной оценке, работа над ним должна была занять около 20 лет. За первые девять лет достигнут очень существенный прогресс, так что проект может завершиться досрочно.

6

Новые решения старой задачи

Гипотеза Морделла

Настало время нам вновь окунуться в теорию чисел и двинуться по направлению к Великой теореме Ферма. Чтобы подготовить почву, я начну с менее известной, но, по мнению некоторых, еще более важной задачи. В 2002 г. Эндрю Гранвиль и Томас Такер представили ее следующим образом:

> «В 1922 г. Морделл написал одну из величайших статей в истории математики… В самом конце статьи он задал пять вопросов, которые сыграли важную роль в мотивировании значительной части исследований XX в. в области диофантовой арифметики. Ответ на самый важный и сложный из этих вопросов дал Фальтингс в 1983 г., выдвинув для этого идеи, которые можно назвать одними из наиболее глубоких и мощных в истории математики».

Упомянутый здесь Луис Морделл — британский специалист по теории чисел, родившийся в США в еврейской семье литовского происхождения, а Герд Фальтингс — немецкий математик. Вопрос, о котором идет речь, приобрел известность как гипотеза Морделла. В цитате, помимо прочего, обозначен ее точный статус: блестяще доказана Фальтингсом.

Гипотеза Морделла относится к крупному отделу теории чисел — к диофантовым уравнениям. Они названы так в честь Диофанта Александрийского, написавшего где-то около 250 г. н. э. знаменитую книгу «Арифметика». Считается, что первоначально она включала в себя 13 книг, но до нас дошли лишь шесть, и все в позднейших копиях. Это не был арифметический текст в буквальном смысле, т. е. речь в нем не шла о сложении и умножении. По существу, это был первый текст по алгебре, в котором были собраны почти все познания греков о том, как нужно решать уравнения. В нем использовалась даже некая рудиментарная форма алгебраического языка: судя по всему, для обозначения неизвестного в ней использовался вариант ς греческой буквы «сигма» (мы для этого используем x), для квадрата неизвестного (вместо нашего x^2) — Δ^Y, а для куба неизвестного (вместо нашего x^3) — K^Y. Сложение обозначалось тем, что символы помещались рядом друг с другом, а вычитание имело собственный символ. Величина, обратная неизвестному (наше $1/x$), выглядела как $\varsigma\chi$ и т. д. Эти обозначения восстановлены на основании позднейших копий и переводов и могут быть не вполне точными. Классическая греческая математика требовала, чтобы решения уравнений были рациональными числами, т. е. дробями вроде 22/7, сформированными из целых чисел. Часто требовалось даже, чтобы они сами были целыми числами. Все задействованные числа были положительными: представление об отрицательных числах появилось несколькими столетиями позже в Китае и Индии. Сегодня мы называем подобные задачи диофантовыми уравнениями. В «Арифметике» можно обнаружить замечательно глубокие результаты. В частности, Диофант, судя по всему, знал, что любое целое число может быть представлено в виде суммы четырех полных квадратов целых чисел (включая нуль). Лагранж впервые доказал это в 1770 г. Но нас в данном случае интересует другой результат — формула для пифагоровых троек, в которых сумма двух полных квадратов дает третий полный квадрат. Название происходит от теоремы Пифагора: именно таким соотношением связаны стороны прямоугольного треугольника. Самый известный пример — знаменитый треугольник 3, 4, 5: ($3^2 + 4^2 = 5^2$). Еще один пример — треугольник 5, 12, 13: ($5^2 + 12^2 = 13^2$). Рецепт поиска пифагоровых троек сформулирован в виде двух лемм (вспомогательных утверждений), помещенных перед Предложениями 29 и 30 в Книге X «Начал» Евклида.

Приведенная у Евклида процедура позволяет получить бесконечно много пифагоровых троек. Морделл знал несколько других диофантовых уравнений, для которых существует формула с бесконечным числом решений. Он знал также, что существует другой тип диофантовых уравнений, имеющих бесконечно много решений, которые *не* описываются формулой. Существуют так называемые эллиптические кривые — достаточно глупое название, поскольку они не имеют практически никакого отношения к эллипсам, — где бесконечность числа решений возникает потому, что любые два решения можно скомбинировать так, чтобы получилось еще одно. Сам Морделл доказал одно из фундаментальных свойств таких уравнений: все бесконечное множество решений может быть получено при помощи этого процесса из конечного их числа.

Помимо этих двух известных типов уравнений, все остальные диофантовы уравнения, которые мог придумать Морделл, попадали в одну из двух категорий. Либо про уравнение было известно, что число его решений конечно (или их просто нет), либо никто не мог сказать наверняка, является ли число его решений конечным или бесконечным. В сущности, ничего нового в этом не было, но Морделлу показалось, что он видит в этом закономерность, которую до него никто не замечал. Закономерность эта относилась вовсе не к теории чисел — скорее, ее можно было отнести к топологии. Чтобы разобраться в этом, необходимо было рассматривать решения уравнений в комплексных числах, а не в рациональных или целых. А это, что ни говори, противоречило самому духу диофантовых уравнений.

Здесь стоит добавить несколько деталей, которые пригодятся нам позже. Не бойтесь формул: они нужны мне в основном для того, чтобы можно было ссылаться на что-то конкретное. Сосредоточьтесь на рассказе, который лежит за ними.

Пифагоровы тройки представляют собой целочисленные решения пифагорова уравнения

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

Разделив обе части уравнения на z^2, получим

$$(x/z)^2 + (y/z)^2 = 1.$$

Согласно главе 3, это означает, что пара рациональных чисел $(x/z, y/z)$ лежит на единичной окружности в плоскости. Далее пифагорово уравнение берет начало в геометрии и имеет геометрическую интерпретацию: связанный с ним треугольник является прямоугольным. Формула, которую я только что вывел, позволяет дать чуть другую геометрическую интерпретацию, причем не одной, а всех пифагоровых троек. Решения пифагорова уравнения непосредственно соответствуют всем рациональным точкам единичной окружности. Мы считаем точку рациональной, если рациональны обе ее координаты.

Из этого можно сделать немало интересных выводов. Если привлечь тригонометрию (но можно обойтись и одной алгеброй), обнаружится, что для любого числа t точка

$$\left(\frac{2t}{t^2+1}, \frac{t^2-1}{t^2+1}\right)$$

лежит на единичной окружности. Более того, если t рационально, то рациональна и эта точка. Все рациональные точки возникают подобным образом, так что мы получили исчерпывающую формулу для всех решений пифагорова уравнения. Она эквивалентна евклидовой формуле, которая, в свою очередь, совпадает с диофантовой. К примеру, если $t = 22/7$, формула даст результат

$$\left(\frac{308}{533}, \frac{435}{533}\right).$$

Можно проверить: $308^2 + 435^2 = 533^2$. Для нас точная формула не слишком важна, важно, что она существует.

Это не единственное диофантово уравнение, для всех решений которого существует единая формула, но таких уравнений относительно немного. Например уравнения Пелля, такие как $x^2 = 2y^2 + 1$. У этого уравнения бесконечно много решений ($3^2 = 2 \times 2^2 + 1$, $17^2 = 2 \times 12^2 + 1$) и для них существует общая формула. Однако упорядоченность пифагоровых троек этим не ограничивается; геометрия подсказывает нам и другие закономерности. Предположим, мы имеем две пифагоровы тройки. Следственно, существует два соответствующих им решения пифагорова уравнения — две рациональные точки на окружности. Геометрия предлагает естественный способ «сложить» эти точки. Начнем с точки (1, 0), в которой окружность пересекает горизонтальную ось, и найдем углы между этой точкой и двумя точками-решениями. Сложим эти два угла (см. рис. 25) и посмотрим, что получится. Точка, разумеется, тоже лежит на окружности, и короткий расчет покажет, что она также рациональна. Таким образом, имея два любых решения, мы можем получить третье. Математики уже заметили множество подобных фактов, причем большинство из них обретает смысл сразу же, как только мы вспоминаем о рациональных точках на окружности.

«Короткий расчет», который я небрежно пропустил, делается с использованием тригонометрии. Классические тригонометрические функции, такие как синус и косинус, теснейшим образом связаны с геометрией окружности. В расчете, на который я ссылался, используются стандартные, довольно элегантные формулы вычисления синуса и косинуса суммы двух углов через синусы и косинусы самих углов. Существует много способов получения синусов и косинусов, и один из них, достаточно изящный, основан на интегральном исчислении. Если вы будете интегрировать алгебраическую функцию $1/\quad 1-x^2$, то результат может быть выражен, используя функцию синус. Точнее, нам нужна функция, обратная синусу: угол, синус которого равен интересующему нас числу[23].

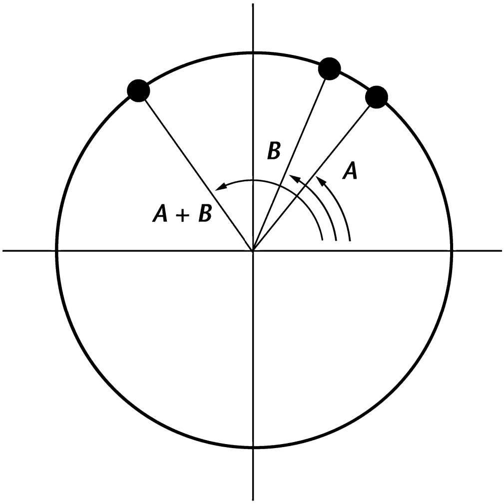


Рис. 25. Сложение двух рациональных решений A и B пифагорова уравнения для получения третьего решения: A + B

Интеграл возникает, когда мы пытаемся вывести формулу длины дуги окружности методами математического анализа, а геометрия окружности дает простое, но очень важное следствие для этого результата. Длина единичной окружности равна 2π, поэтому пройдя расстояние 2π вдоль окружности, вы окажетесь в точности на том же месте. То же можно сказать о любом расстоянии, кратном 2π: по стандартному математическому соглашению положительные целые числа соответствуют направлению против часовой стрелки, а отрицательные — по часовой стрелке. Следовательно, синус и косинус числа остаются неизменными при добавлении к аргументу величины 2π, взятой целое число раз. Мы говорим, что эти функции периодические с периодом 2π.

Аналитики XVIII и XIX вв. обобщили эту интегральную формулу и нашли целую группу интересных новых функций, аналогичных знакомым тригонометрическим. Эти новые функции выглядели загадочно; они были периодическими, как синус и косинус, но хитроумно периодическими. Вместо одного периода, к примеру 2π (или кратных ему), они имели два независимых периода. Если вы попытаетесь проделать такое с действительными функциями, то получите всего лишь константы, но для комплексных чисел здесь открываются широкие возможности.

Начало исследованиям в этой области положили итальянский математик Джулио ди Фаньяно и плодовитый Эйлер. Фаньяно пытался при помощи интегрального исчисления найти длину дуги эллипса, но не сумел вывести формулу в явном виде. Сегодня это уже не удивительно, ведь мы знаем, что такой формулы не существует. Однако он заметил, что длины различных особых дуг эллипса находятся между собой в определенных отношениях. Результаты своих исследований Фаньяно опубликовал в 1750 г. Эйлер в аналогичной ситуации заметил те же отношения и представил их в виде формального отношения интегралов. Они похожи на интеграл, связанный с функцией синуса, но квадратичное выражение под знаком квадратного корня сменяется кубическим многочленом или многочленом четвертой степени, к примеру, таким: $(1 - x^2)\ (1 - 4x^2)$.

В 1811 г. Адриен Мари Лежандр опубликовал первую книгу объемного трехтомного трактата об этих интегралах, известных как эллиптические благодаря своей связи с длиной дуги сегмента эллипса. Однако он умудрился пройти мимо самого важного их свойства: существования новых функций, аналогичных синусу и косинусу, *обратные* функции к которым достаточно просто выражают величину интеграла[24]. Гаусс, Нильс Хенрик Абель и Карл Якоби быстро заметили упущение. Гаусс оставил свои мысли при себе, что для него вполне типично.

Абель в 1826 г. представил во Французскую академию собственную работу, но Коши, президент Академии, потерял рукопись, и опубликована она была только в 1841 г., через 12 лет после трагической ранней кончины Абеля от чахотки. Однако в 1827 г. вышла другая статья Абеля на ту же тему. Якоби положил новые «эллиптические функции» в основу громадного тома, опубликованного в 1829 г., и этот труд направил исследования в области комплексного анализа по совершенно новому пути.

В результате был выявлен целый комплекс взаимосвязанных свойств, аналогичных свойствам тригонометрических функций. Оказалось, что соотношение, замеченное Фаньяно и Эйлером, можно интерпретировать иначе, в виде простого списка формул, связывающих эллиптические функции суммы двух чисел с эллиптическими функциями самих чисел. Но самая интересная черта эллиптических функций оставляет тригонометрические функции далеко позади. Эллиптические функции не просто периодические; для них характерна двойная периодичность. Линия одномерна, поэтому и рисунок ее может повторяться только в одном направлении, вдоль линии. Комплексная плоскость двумерна, так что рисунок на ней может повторяться, как на обоях: вдоль бумажной полосы и одновременно поперек — вдоль стены, на соседние полосы. С каждой эллиптической функцией связаны два независимых комплексных числа (ее периоды), прибавление любого из которых к переменной не меняет значение функции.

Повторяя этот процесс, мы приходим к выводу, что значение функции не меняется при добавлении к переменной любой целочисленной комбинации двух периодов. Эти комбинации можно интерпретировать и геометрически: они определяют на комплексной плоскости решетку, которая разбивает плоскость на параллелограммы; все, что происходит в одном параллелограмме, повторяется и во всех остальных (см. рис. 26). Если мы рассмотрим отдельно взятый параллелограмм и то, как он соединяется с соседними, то получим, что нам придется отождествить противоположные его стороны, как тор определяется через отождествление противоположных сторон квадрата (см. рис. 12). Параллелограмм, противоположные стороны которого попарно отождествляются, топологически тоже представляет собой тор. Иными словами, точно так же, как синус и косинус связаны с окружностью, эллиптические функции связаны с тором.

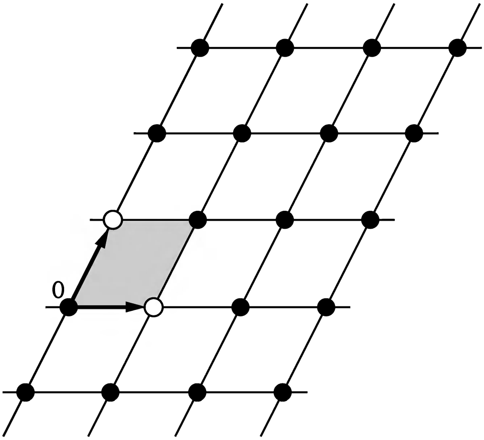


Рис. 26. Решетка на комплексной плоскости. Стрелки указывают на два периода, обозначенные белыми точками. Значение функции в затемненном параллелограмме определяет ее значение и во всех остальных параллелограммах

Есть здесь и связь с теорией чисел. Я сказал, что функция, обратная синусу, получается путем интегрирования формулы, в которой фигурирует квадратный корень из квадратного трехчлена (т. е. многочлена второй степени). Эллиптические функции в этом похожи, но квадратный трехчлен в них заменяется на многочлен третьей или четвертой степени. Случай с четвертой степенью уже упоминался, потому что исторически он появился первым, но теперь давайте сосредоточимся на случае с многочленом третьей степени. Если мы обозначим квадратный корень как y, а многочлен как $ax^3 + bx^2 + cx + d$, где

a, b, c и d — числовые коэффициенты, то x и y удовлетворяют уравнению:

$$y^2 = ax^3 + bx^2 + cx + d.$$

Это уравнение можно рассматривать в нескольких различных контекстах, в зависимости от того, какие ограничения наложены на переменные и коэффициенты. Если все это действительные числа, уравнение определяет кривую на плоскости. Если это комплексные числа, то специалисты по алгебраической геометрии все равно называют множество решений этого уравнения кривой просто по аналогии. Но теперь это кривая в пространстве пар комплексных чисел, четырехмерном в действительных координатах. И кривая в данном случае — это поверхность с точки зрения действительных координат.

На рис. 27 показаны типичные действительные эллиптические кривые $y^2 = 4x^3 - 3x + 2$ и $y^2 = 4x^3 - 3x$. Поскольку y появляется в уравнении в виде квадрата, кривая симметрична относительно горизонтальной оси. В зависимости от коэффициентов это либо одна волнообразная кривая, либо кривая с отдельным овальным компонентом. В комплексных числах кривая всегда представляет собой единое связное множество.

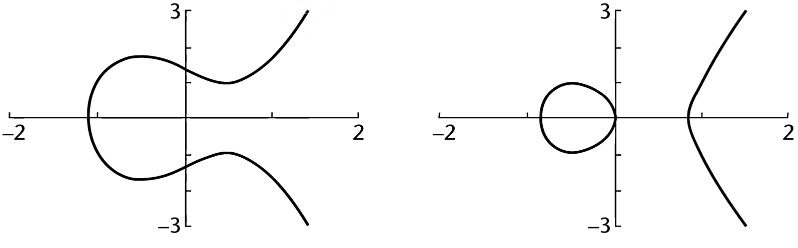


Рис. 27. Типичные действительные эллиптические кривые: *слева* — $y^2 = 4x^3 - 3x + 2$; *справа* — $y^2 = 4x^3 - 3x$

Если мы потребуем, чтобы переменные и коэффициенты были рациональными, в игру вступит теория чисел и мы получим диофантово уравнение. Его графическое представление зачем-то называется эллиптической кривой, хотя совершенно не похоже на эллипс. Все дело в том, что это уравнение связано с эллиптическими функциями. Это как назвать окружность треугольной кривой только потому, что она связана с тригонометрией. Однако это название уже высечено на скрижалях, так что нам придется мириться с ним.

Теория эллиптических функций — глубокая и богатая теория, математики открыли у эллиптических кривых бессчетное количество красивых свойств. Одно из них аналогично тому, как мы объединяем два решения пифагорова уравнения, складывая соответствующие углы и получая третье решение. Две точки на эллиптической кривой можно «сложить», проведя через них прямую линию и посмотрев, в какой точке она пересечет кривую в третий раз (см. рис. 28). (Заметим, что третья точка обязательно существует, поскольку уравнение кубическое. Однако она может оказаться «в бесконечности» или совпасть с одной из первых двух точек, если прямая пройдет по касательной к кривой.) Если первые две точки у нас обозначены как *P* и *Q*, обозначим третью как *P***Q*.

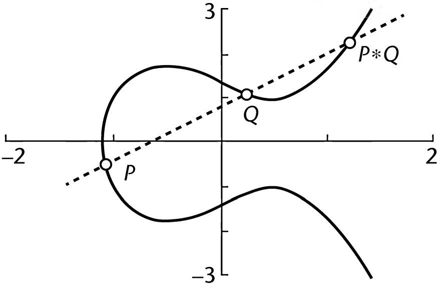


Рис. 28. Сложение точек *P* и *Q* с получением точки *P***Q*

Расчет показывает, что если *P* и *Q* — рациональные точки, то и точка *P***Q* рациональна. Операция * придает набору рациональных точек алгебраическую структуру, но оказывается полезной и при рассмотрении еще одной связанной с этим операции. Выберем любую рациональную точку O на кривой и определим:

$$P + Q = (P^*Q)\ ^*O.$$

Эта новая операция подчиняется некоторым фундаментальным законам обычной алгебры, причем O ведет себя как нуль и превращает множество всех рациональных точек в то, что специалисты по алгебре называют группой (см. главу 10). Важно, что здесь, как с пифагоровыми тройками, можно «сложить» любые два решения и получить третье. То, что «групповой закон» действует на рациональные точки, поразительно: в частности, это означает, что стоит нам найти два рациональных решения диофантова уравнения, и мы автоматически получим множество других решений.

Около 1908 г. Пуанкаре задался вопросом: существует ли конечный набор решений, такой, чтобы из него можно было получить все остальные решения последовательным применением групповой операции? Это важно, потому что из существования такого набора следует, что *все* рациональные решения можно охарактеризовать при помощи конечного списка решений. В интереснейшей работе 1922 г. Морделл доказал, что ответ на вопрос Пуанкаре положителен. После этого эллиптические кривые стали центральным элементом теории чисел, поскольку далеко не все диофантовы уравнения могут похвастать такой степенью контроля. Итак, эллиптические кривые, как и пифагорово уравнение, имеют бесконечное множество рациональных решений. Многие диофантовы уравнения, напротив, имеют конечное число или вообще не имеют решений. Я собираюсь немного отклониться от темы и поговорить о целом семействе подобных уравнений и полученном недавно замечательном доказательстве того, что существуют лишь очевидные решения.

Пифагорейцы были увлечены своим уравнением, потому что верили: в основе Вселенной лежат числа. В поддержку этой философии они обнаружили, что музыкальной гармонией управляют простые числовые отношения. Это было установлено экспериментально при помощи наблюдений за звучанием натянутой струны. Струна с такой же степенью натяжения, но вдвое меньшей длины, дает ноту на октаву выше. Это самое гармоничное сочетание двух нот — настолько гармоничное, что звучит даже несколько простовато. В западной музыке следующая по значению гармония — кварта, где одна струна по длине составляет 3/4 другой струны, и квинта, где длина одной струны составляет 2/3 от длины другой.

Начав с 1 и умножая последовательно на 2 или 3, можно получить числа 2, 3, 4, 6, 8, 9, 12 и т. д. — числа вида 2^a3^b. Благодаря связи с музыкой они получили название гармонических. В XIII в. во Франции еврейский ученый по имени Гершон бен Соломон Каталан написал книгу «Врата небес», посвятив три ее части физике, астрономии и метафизике. В 1343 г. епископ Мо убедил его сына (по крайней мере историки считают, что, вероятно, это был его сын) Леви бен Гершома написать математическую книгу «Гармония чисел». В нее вошла, в частности, задача, которую впервые поставил композитор и теоретик музыки Филипп де Витри: когда два гармонических числа могут различаться на единицу? Подобные пары найти несложно: сам де Витри знал их четыре: (1, 2), (2, 3), (3, 4) и (8, 9). Но Гершом доказал, что это все возможные решения и других не существует.

Среди перечисленных де Витри пар гармонических чисел наибольший интерес привлекает пара (8, 9). Первое число в ней — куб, 2^3; второе — квадрат, 3^2. Математики заинтересовались, могут ли другие квадраты и кубы различаться на единицу; Эйлер доказал, что не могут, за исключением тривиального случая (0, 1) и случая (–1, 0), если разрешены отрицательные числа. В 1844 г. уже другой Каталан опубликовал более всеобъемлющее заявление, о котором, должно быть, думали многие математики, но которое никто до него не счел нужным озвучить. Речь идет о бельгийском математике Эжене Шарле

Каталане, который в 1844 г. написал в один из ведущих математических журналов того времени *Journal für die Riene und Angewandte Mathematik* следующее:

> «Покорнейше прошу вас, сударь, объявить в вашем журнале следующую теорему, кою я почитаю верной, хотя до сего момента и не преуспел в доказательстве; быть может, другие добьются большего успеха. Два последовательных целых числа, кроме 8 и 9, не могут быть последовательными степенями; иначе говоря, уравнение $x^m - y^n = 1$, в котором все неизвестные — положительные целые числа, допускает лишь одно решение».

Это утверждение получило известность как гипотеза Каталана. Показатели степени m и n здесь — целые числа больше 1.

Несмотря на частичные успехи, гипотеза Каталана долгое время упорно противостояла всем усилиям математиков, пока в 2002 г. ее не доказал вдруг Преда Михайлеску. Этот математик родился в 1955 г. в Румынии, а в 1973 г. поселился в Швейцарии и к тому времени только что получил докторскую степень. Темой его диссертации было «Деление кругов и проверка на простоту», и речь в ней шла о применении теории чисел при проверке на простоту (глава 2). Эта задача не имеет особого отношения к гипотезе Каталана, но Михайлеску пришло в голову, что методы-то его точно имеют к ней отношение. В них он отталкивался от идей, уже упоминавшихся мной в главе 3: гауссового построения правильного 17-угольника и связанных с этим алгебраических уравнений, решения которых называют круговыми числами. Доказательство получилось достаточно сложным и произвело в математическом сообществе настоящий фурор. Из него явствует, что какие бы величины мы ни выбрали для двух показателей степени, число решений остается конечным — и помимо очевидных решений с участием 0 и ±1 единственное решение, представляющее интерес, это $3^2 - 2^3 = 1$.

Приведенные примеры наглядно показывают, что одни диофантовы уравнения имеют бесконечное множество решений, а другие — нет. Что тут интересного, скажете вы, эти два варианта перекрывают весь спектр возможностей. Но если вы спросите, какие уравнения к какому типу относятся, все станет намного интереснее.

Морделл готовил учебник по диофантовым уравнениям. В то время эта область математики напоминала биологию на раннем этапе ее развития: множество собранных коллекционерами бабочек, и никакой систематической классификации. Наука пребывала почти в том же состоянии, в каком ее оставил Диофант, только беспорядка стало еще больше: это был бессистемный набор ловких трюков, для каждого типа уравнений свой, и вот такой не слишком подходящий для учебника материал отчаянно нуждался в систематизации, чем Морделл и занялся.

В какой-то момент он, должно быть, заметил, что все уравнения, достоверно имеющие бесконечное множество рациональных решений, — такие как пифагорово уравнение или эллиптические кривые — имеют одну общую черту. Он сосредоточился на одном классе уравнений — на тех, в которых (после перевода в рациональную форму, как я сделал с пифагоровым уравнением) присутствует всего две переменных. Есть два случая, когда нам заранее известно, как найти бесконечное множество решений. Пример одного из них — пифагорово уравнение в эквивалентной форме $x^2 + y^2 = 1$. В этом случае существует формула для нахождения решений. Вставьте в эту формулу любое рациональное число и получите рациональное решение, причем формула позволяет получить все решения. Пример второго случая — эллиптические кривые: здесь существует *процесс,* позволяющий получать новые решения из старых, и гарантия того, что если начать с подходящего конечного набора решений, этот процесс позволит получить их все.

Гипотеза Морделла утверждает, что во всех случаях, когда уравнение имеет бесконечное множество рациональных решений, должно присутствовать одно из названных свойств. Существует либо общая формула, либо процесс, позволяющий получить все решения из подходящего конечного их набора. Во всех остальных случаях число рациональных решений конечно; примером могут служить уравнения вида $x^m - y^n = 1$, которые фигурируют в гипотезе Каталана. Решения здесь в каком-то смысле случайны и не имеют никакой структурной основы.

Морделл пришел к этому наблюдению немного иным путем. Он обратил внимание на то, что все уравнения с бесконечным числом рациональных решений имеют одно поразительное топологическое свойство. У всех у них род равен 0 или 1. Вспомните из главы 4, что род — это понятие из области топологии кривых, которое указывает на то, сколько в данной поверхности отверстий. Род сферы равен 0, род тора — 1, род тора с двумя отверстиями равен 2 и т. д. Но откуда берутся поверхности в задаче из теории чисел? Из координатной геометрии. Мы видели, что пифагорово уравнение, которое интерпретировано в терминах рациональных чисел и допускает действительные решения, определяет окружность. Морделл сделал еще один шаг и допустил комплексные решения. Любое уравнение с двумя комплексными переменными определяет то, что специалисты по алгебраической геометрии называют комплексной кривой. Однако с точки зрения действительных чисел и зрительного восприятия человека каждое комплексное число двумерно: у него на деле две компоненты — действительная и мнимая части. Так что «кривая» в комплексном смысле есть поверхность для нас с вами. И, как у всякой поверхности, у нее есть род, — вот и все.

Всякий раз, когда про уравнение было известно, что оно имеет конечное число решений, его род равнялся по крайней мере двум. Род важных уравнений, статус которых относительно числа решений не был известен, тоже равнялся по крайней мере двум. Морделл на основании достаточно хлипких, как тогда казалось, указаний сделал отчаянно смелый шаг: он предположил, что любое диофантово уравнение с родом 2 или больше имеет конечное число рациональных решений. Так, по мановению его руки, диофантовы бабочки аккуратно распределились по родственным семействам; точнее говоря, по родам (даже термин подходящий).

В гипотезе Морделла была всего лишь одна крохотная загвоздка. Она связала между собой две чрезвычайно разные вещи: рациональные решения и топологию. В то время подобная связь казалась в высшей степени неубедительной. Если даже она существовала, то никто не знал, как ее искать; непонятно было даже, как подступиться к этой проблеме. Так что гипотеза представляла собой отчаянное, ничем не подтвержденное заявление, обещавшее, однако, громадные потенциальные дивиденды.

В 1983 г. Фальтингс опубликовал эффектное доказательство того, что фантастическое предположение Морделла на самом деле верно. Его доказательство было построено на методах алгебраической геометрии. Совершенно другое доказательство, основанное на аппроксимации действительных чисел рациональными, вскоре нашел Пауль Войта, а в 1990 г. Энрико Бомбиери опубликовал упрощенное доказательство, основанное примерно на тех же принципах. Существует приложение теоремы Фальтингса к Великой теореме Ферма — проблеме, о которой мы будем подробно говорить в главе 7. Оно утверждает, что для любого целого n, большего или равного 3, уравнение $x^n + y^n = 1$ имеет конечное число целых решений. Род соответствующей кривой равен $(n-1)(n-2)/2$, а это по крайней мере 3, если $n \geq 4$. Теорема Фальтингса прямо подразумевает, что при любом $n \geq 4$ уравнение Ферма имеет в лучшем случае конечное число рациональных решений. Ферма утверждал, что оно вовсе не имеет решений, за исключением случаев, когда x или y равны нулю, так что это было очень серьезное продвижение. В следующей главе мы вернемся к истории Великой теоремы Ферма и посмотрим, как заявление ученого подтвердилось.

7

«Недостаточные» поля

Великая теорема Ферма

Впервые мы столкнулись с Ферма в главе 2, где его элегантная теорема о степенях чисел обеспечила метод проверки чисел на простоту. Эта глава посвящена куда более сложному утверждению: Великой теореме Ферма. Название звучит загадочно. «Теорема» — это вроде бы понятно, но кто такой был Ферма и почему эту теорему называют *великой,* а иногда *последней* его теоремой? Может быть, название — всего лишь хитрый маркетинговый ход? Оказывается, нет: такое название эта задача получила в XVIII в., когда лишь несколько ведущих математиков хотя бы слышали о ней и уж тем более интересовались ею. Но Великая, или Последняя, теорема Ферма и вправду загадочна.

Пьер Ферма родился во Франции в 1601 г. по одним источникам и в 1607–1608 гг. по другим. Не исключено, что путаница возникла из-за его брата, носившего такое же имя. Его отец был зажиточным купцом, он торговал кожей и занимал высокое положение в городе, а мать происходила из семьи юристов. Пьер учился в университете в Тулузе, а в конце 1620-х гг. перебрался в Бордо, где у него проявился талант к математике. Он говорил на нескольких языках. Кроме того, он собирал материалы и работал над восстановлением одного из классических древнегреческих трудов по математике, принадлежавшего перу Аполлония и давно утраченного. Своими многочисленными открытиями Ферма делился с ведущими математиками своего времени.

В 1631 г. Ферма получил ученую степень юриста в университете Орлеана и был назначен советником в суд Тулузы. Это назначение принесло ему дворянский титул и дало право добавлять к фамилии частицу «де»: де Ферма. Советником суда и практикующим юристом он оставался до конца жизни. Однако страстью его была математика. Он почти ничего не публиковал, предпочитая излагать свои открытия в письмах к коллегам-математикам, как правило, без доказательств. В математике Ферма так и остался любителем, но его работы пользовались заслуженным признанием профессионалов, со многими из которых он был знаком достаточно близко, хотя и по переписке. По существу, он и был профессионалом; просто не занимал в математике никакого официального поста.

Некоторые из его доказательств дошли до нас в письмах и заметках; ясно, что Ферма прекрасно представлял себе, что такое настоящее доказательство. После его смерти многие из его наиболее глубоких теорем остались недоказанными, и за них взялись профессионалы. Через несколько десятилетий лишь одному из утверждений Ферма по-прежнему недоставало доказательства; естественно, именно это утверждение получило известность как его последняя теорема. В отличие от остальных, она никак не поддавалась усилиям математиков и вскоре прославилась контрастом между простотой формулировки и очевидной сложностью поиска доказательства.

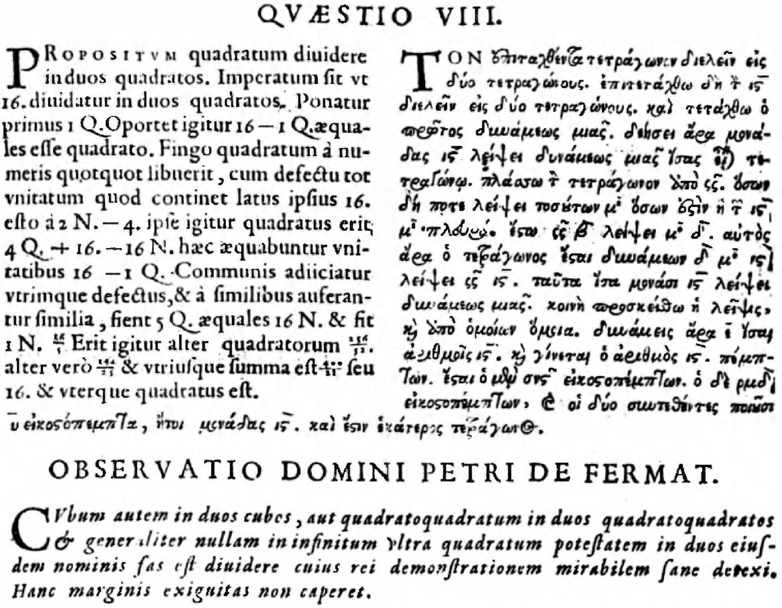
Судя по всему, Ферма пришел к своей знаменитой теореме около 1630 г. Точная дата неизвестна, но произошло это вскоре после того, как он начал читать недавно изданную «Арифметику» Диофанта. Тогда у него и появилась идея этой теоремы. Опубликована она была впервые в 1670 г., через пять лет после смерти Ферма. Его сын Самюэль выпустил необычное издание «Арифметики» Диофанта, которое включало и заметки на полях, сделанные Пьером Ферма в его личном экземпляре латинского перевода. Этот перевод был сделан Клодом Гаспаром Баше де Мезириаком и издан в 1621 г. Великая теорема Ферма изложена там в виде заметки к диофантову вопросу VIII Книги II (см. рис. 29).

Речь в этом месте книги шла о задаче представления полного квадрата как суммы двух полных квадратов. Из главы 6 мы знаем, что таких пифагоровых троек существует бесконечное множество. Диофант задается тем же вопросом, но в несколько более сложной формулировке: как найти две меньшие стороны пифагорова треугольника, если известна самая большая его сторона. Иными словами, конкретный квадрат следует «разложить» на два квадрата и выразить в виде их суммы. Он показывает, как решить эту задачу, если бо́льшая сторона треугольника равна 4, и получает ответ

$$4^2 = (16/5)^2 + (12/5)^2$$

в рациональных числах. Умножив все на 25, получим $20^2 = 16^2 + 12^2$, а поделив затем на 16, получим знакомое $3^2 + 4^2 = 5^2$. Диофант обычно иллюстрировал общие методы конкретными примерами и не приводил никаких доказательств; такая традиция восходит еще к Древнему Вавилону.

Экземпляр «Арифметики» с собственноручными заметками Ферма не сохранился, но, должно быть, такая запись в нем была, поскольку Самюэль прямо об этом говорит. Вряд ли Ферма стал бы долго таить такое сокровище, да и само предположение настолько естественно, что мысль о нем, вероятно, пришла Пьеру в голову сразу же по прочтении восьмого вопроса второй книги «Арифметики». Очевидно, ему стало интересно, можно ли проделать что-нибудь подобное с кубами вместо квадратов — согласитесь, естественный для математика вопрос. Он не нашел подобных примеров — мы можем быть в этом уверены, поскольку точно знаем, что их не существует; неудача ждала его и в случае с более высокими степенями, к примеру с четвертой. Он решил, что эти задачи не имеют



QVÆSTIO VIII.

PROPOSITVM quadratum diuidere in duos quadratos. Imperatum ſit vt 16. diuidatur in duos quadratos. Ponatur primus 1 Q. Oportet igitur 16 — 1 Q. æquales eſſe quadrato. Fingo quadratum à numeris quotquot libuerit, cum defectu tot vnitatum quod continet latus ipſius 16. eſto à 2 N. — 4. ipſe igitur quadratus erit 4 Q. + 16. — 16 N. hæc æquabuntur vnitatibus 16 — 1 Q. Communis adiiciatur vtrimque defectus, & à ſimilibus auferantur ſimilia, fient 5 Q. æquales 16 N. & fit 1 N. $\frac{16}{5}$ Erit igitur alter quadratorum $\frac{256}{25}$. alter verò $\frac{144}{25}$ & vtriuſque ſumma eſt $\frac{400}{25}$ ſeu 16. & vterque quadratus eſt.

ΤΟΝ ἐπιταχθέντα τετράγωνον διελεῖν εἰς δύο τετραγώνους. ἐπιτετάχθω δὴ τὸν ις̅ διελεῖν εἰς δύο τετραγώνους. καὶ τετάχθω ὁ πρῶτος δυνάμεως μιᾶς. δεήσει ἄρα μονάδας ις̅ λείψει δυνάμεως μιᾶς ἴσας εἶναι τετραγώνῳ. πλάσσω τὸν τετράγωνον ἀπὸ ςς̅. ὅσων δή ποτε λείψει τοσούτων μο ὅσων ἐστὶν ἡ τῶν ις̅ μο πλευρά. ἔστω ςς̅ β̅ λείψει μο δ̅. αὐτὸς ἄρα ὁ τετράγωνος ἔσται δυνάμεων δ̅ μο ις̅ λείψει ςς̅ ις̅. ταῦτα ἴσα μονάσι ις̅ λείψει δυνάμεως μιᾶς. κοινὴ προσκείσθω ἡ λεῖψις, καὶ ἀπὸ ὁμοίων ὅμοια. δυνάμεις ἄρα ε̅ ἴσαι ἀριθμοῖς ις̅. καὶ γίνεται ὁ ἀριθμὸς ις̅. πέμπτων. ἔσται ὁ μὲν σνς̅ εἰκοσοπέμπτων. ὁ δὲ ρμδ̅ εἰκοσοπέμπτων, καὶ οἱ δύο συντεθέντες ποιοῦσι υ̅ εἰκοσόπεμπτα, ἤτοι μονάδας ις̅. καὶ ἔστιν ἑκάτερος τετράγωνος.

OBSERVATIO DOMINI PETRI DE FERMAT.

Cubum autem in duos cubos, aut quadratoquadratum in duos quadratoquadratos & generaliter nullam in infinitum vltra quadratum poteſtatem in duos eiuſdem nominis fas eſt diuidere cuius rei demonſtrationem mirabilem ſane detexi. Hanc marginis exiguitas non caperet.

Рис. 29. Заметка Пьера Ферма на полях книги, опубликованная в изданной его сыном «Арифметике» Диофанта

решений. Заметка на полях говорит именно об этом. В переводе это звучит примерно так:

> «Невозможно разложить куб на два куба, биквадрат на два биквадрата и вообще никакую степень, большую квадрата, на две степени с тем же показателем. Я нашел этому поистине чудесное доказательство, но поля книги слишком узки для него».

Говоря алгебраическим языком, Ферма, согласно его собственному заявлению, доказал, что диофантово уравнение

$$x^n + y^n = z^n$$

не имеет целочисленных решений, если n — любое целое число, большее или равное 3. Ясно, что при этом он не рассматривал тривиальных решений, при которых *x* или *y* равны нулю. Чтобы не повторять это уравнение постоянно, я буду называть его в дальнейшем уравнением Ферма.

Если у Ферма действительно было доказательство, то найти его так никому и не удалось. В конце концов теорема была доказана в 1995 г., больше чем через три с половиной столетия после появления, но методы доказательства выходят далеко за рамки методик, доступных во времена Ферма или даже таких, которые он мог бы сам изобрести. Надо сказать, что поиски доказательства этой теоремы оказали громадное влияние на развитие математики. По существу, именно они привели к созданию алгебраической теории чисел, которая расцвела в XIX в. благодаря очередной неудачной попытке доказать теорему и блестящей идее, которая едва не спасла доказательство. В конце XX — начале XXI в. она дала толчок настоящей революции.

Вначале математики, работавшие над Великой теоремой Ферма, пытались перебирать степени одну за другой. Общего доказательства теоремы, о котором говорил ее автор в заметке на полях, могло и не быть, но нам известно, как Ферма доказал свою теорему для четвертых степеней. Главный инструмент здесь — евклидова методика поиска пифагоровых троек. Четвертая степень числа — это квадрат квадрата этого числа. Так что любое решение уравнения Ферма для четвертых степеней — это пифагоров треугольник, в котором все три числа также являются полными квадратами. Это дополнительное условие можно ввести в методику Евклида и после некоторых хитрых маневров получить *еще одно* решение уравнения Ферма для четвертых степеней. Может показаться, что в этом нет никакого особого прогресса; после страницы алгебраических вычислений задача сводится к первоначальной. Однако на самом деле это нам поможет: числа во втором решении меньше, чем в первом (гипотетическом). Главное, если первое решение нетривиально (т. е. если *x* и *y* в нем не равны нулю), то же можно сказать и о втором решении. Ферма указывал, что повторение этой процедуры даст нам последовательность решений, в которой числа становятся все меньше и меньше. Однако любая убывающая последовательность целых чисел должна когда-нибудь остановиться. Это логическое противоречие, так что гипотетического решения, с которого все началось, не существует. Ферма назвал этот метод доказательства методом «бесконечного спуска». Мы сегодня назвали бы его доказательством по методу математической индукции, упомянутому в главе 4. Его, кстати, тоже можно переформулировать в терминах минимальных контрпримеров, или в данном случае минимальных положительных примеров. Предположим, существует положительный пример — нетривиальное решение нашего уравнения. Тогда существует и минимальный положительный пример. Но, согласно рассуждениям Ферма, это означает, что существует еще меньший пример, а это уже противоречие. Следовательно, положительных примеров не существует. Со времен Ферма появились и другие доказательства теоремы для четвертых степеней, и на сегодняшний день их известно около 30.

Ферма использовал тот простой факт, что четвертая степень — это особый случай квадрата. Та же идея показывает, что в целях доказательства теоремы Ферма можно считать, что показатель степени n либо равен 4, либо является нечетным простым числом. Любое число n больше 2 делится либо на 4, либо на некоторое нечетное простое p, так что любая n-я степень — это одновременно либо 4-я степень, либо p-я[1]. За два столетия после Ферма его Великую теорему удалось доказать ровно для трех нечетных простых чисел: это 3, 5 и 7. С кубами разобрался Эйлер в 1770 г.; в его опубликованном доказательстве есть пробел, но его можно заполнить при помощи

---

[1] Имеется в виду, что если теорема Ферма доказана для показателя m, то она автоматически доказана и для любого показателя, кратного m. Таким образом, требуется доказать ее только для простых степеней, начиная от 3, и отдельно для показателя 4, поскольку для 2 она неверна. — *Прим. пер.*

результата, опубликованного им же в другом месте. С пятыми степенями справились Лежандр и Петер Лежен Дирихле около 1825 г. Теорему Ферма для седьмых степеней доказал Габриель Ламе в 1839 г. Позже для этих случаев было найдено немало других доказательств. Где-то по пути несколько математиков получили доказательства для степеней 6, 10 и 14, но эти результаты перекрывались доказательствами для 3, 5 и 7.

Каждое из упомянутых доказательств использует какие-то алгебраические черты, присущие именно этим степеням. Долгое время не было никаких намеков на какую бы то ни было общую структуру, которая могла бы послужить основой доказательства теоремы для всех или хотя бы для значительного числа разных степеней. С ростом показателей степени доказательства становились все сложнее и сложнее. Требовались свежие идеи, открывающие новые горизонты. Софи Жермен, одна из величайших женщин-математиков, разделила теорему Ферма для простых степеней p на два случая. В первом случае ни одно из чисел x, y, z не делится на p. Во втором — одно из них делится. Рассмотрев особые «вспомогательные» простые числа, связанные с p, она доказала, что в первом случае уравнение Ферма не имеет решений для нечетных простых чисел меньших 100. Однако трудно было доказать что-нибудь насчет вспомогательных простых чисел в целом.

Жермен переписывалась с Гауссом, причем сначала под мужским псевдонимом, и оригинальность ее рассуждений весьма впечатлила великого математика. Когда же выяснилось, что его корреспондент — женщина, Гаусс впечатлился еще сильнее и прямо сказал об этом. В отличие от многих своих современников, Гаусс не считал женщин неспособными к высокоинтеллектуальной деятельности, в частности к математическим исследованиям. Позже Жермен предприняла неудачную попытку доказать первый случай Великой теоремы Ферма для всех четных чисел, где опять же можно было бы воспользоваться евклидовой характеристикой пифагоровых троек. Окончательно разобраться с четными степенями удалось только Гаю Тержаняну в 1977 г. Второй случай казался куда более крепким орешком, и никто особенно далеко с ним и не продвинулся.

В 1847 г. Ламе, опираясь на свое доказательство для седьмых степеней, выдвинул замечательную идею. Для ее реализации требовалось ввести комплексные числа, но к тому моменту это уже никого не смущало. Главным ингредиентом было то же, чем воспользовался Гаусс при построении своего правильного 17-угольника (см. главу 3). Любой специалист по теории чисел знал об этом, но до Ламе никому не приходило в голову, что этим можно воспользоваться для доказательства Великой теоремы Ферма.

В системе действительных чисел единица имеет ровно один корень p-й степени (если p нечетное), и корень этот равен самой единице. Но в комплексных числах 1 имеет несколько, а именно p, корней p-й степени. Этот факт — следствие основополагающей теоремы алгебры, поскольку эти корни удовлетворяют уравнению $x^p - 1 = 0$ степени p. Для комплексных корней p-й степени из единицы существует симпатичная формула, из которой явствует, что все они являются степенями $1, \zeta, \zeta^2, \zeta^3, \ldots, \zeta^{p-1}$ некоего комплексного числа ζ (см. прим. [25]). Определяющее свойство этих чисел подразумевает, что $x^p + y^p$ раскладывается на p множителей:

$$x^p + y^p = (x + y)\,(x + \zeta y)\,(x + \zeta^2 y)\,\ldots\,(x + \zeta^{p-1} y).$$

Согласно уравнению Ферма, это выражение равно также z^p, что представляет собой p-ю степень некоего целого числа. Несложно заметить, что если произведение чисел, не имеющих общих делителей, представляет собой p-ю степень, то и каждое число в отдельности представляет собой p-ю степень. Таким образом, если оставить в стороне некоторые технические подробности, Ламе мог записать каждый из сомножителей как p-ю степень. Отсюда он вывел противоречие.

В марте 1847 г. Ламе выступил с полученным в результате доказательством теоремы Ферма в Парижской академии и сказал, что основной идеей он обязан Жозефу Лиувиллю. Лиувилль поблагодарил Ламе, но одновременно указал на потенциальную проблему в доказательстве. Дело в том, что главное утверждение о том, что каждый сомножитель представляет собой p-ю степень, вовсе не бесспорно. Все зависит от единственности разложения на простые множители — причем не для обычных целых чисел, где это свойство выполняется, но для новых типов чисел, введенных Ламе. Эти комбинации степеней ζ известны как круговые, или циклотомические, числа. Слово «циклотомический» означает «разрезающий круг» и указывает на обстоятельство, которое исследовал еще Гаусс. Мало того, что свойство единственности разложения на простые множители для круговых чисел не доказано, сказал Лиувилль, они вполне могут им и не обладать.

У других математиков сомнения возникли даже раньше. За три года до этого Готтхольд Эйзенштейн писал одному из коллег:

> «Если бы у кого-то была теорема, которая утверждала бы, что произведение двух комплексных чисел может делиться на простое число, только если на него делится один из множителей, — что кажется совершенно очевидным, — то он получил бы целую теорию [алгебраических чисел] разом; но такая теорема совершенно неверна».

Теорема, о которой идет речь, есть главный шаг, необходимый для доказательства единственности разложения на простые множители. Эйзенштейн говорил не только о числах, нужных Ламе, но и об аналогичных числах, возникающих при решении других уравнений. Они называются алгебраическими числами. Алгебраическое число — это комплексное число, удовлетворяющее полиномиальному уравнению с рациональными коэффициентами. Алгебраическое целое число — это комплексное число, удовлетворяющее полиномиальному уравнению с целыми коэффициентами, если коэффициент при наибольшей степени x равен 1. Для каждого такого полинома мы получаем связанное с ним поле алгебраических чисел (это означает, что можно складывать, вычитать, умножать и делить такие числа, получая при этом числа того же рода) и соответствующее кольцо (что означает то же самое, за исключением деления) алгебраических целых чисел. Это основные объекты изучения алгебраической теории чисел.

Если, к примеру, взять многочлен $x^2 - 2$, то у него есть корень. Поле включает в себя все числа $a + b$, где a, b — рациональные числа; кольцо целых чисел состоит из чисел такого же вида, где a, b — целые. Здесь опять же простые делители могут быть определены, и притом единственным образом. Но есть и сюрпризы: у многочлена $x^2 + x - 1$ есть корень $(\quad 5 \; - 1)/2$, так что, несмотря на дробь, это алгебраическое *целое* число.

В алгебраической теории чисел сложность заключается не в том, чтобы найти множители. К примеру, круговое число является делителем другого кругового числа, если второе число можно получить умножением первого на еще какое-нибудь круговое число. Определить простые числа также не сложно: круговое целое число является простым, если у него нет других делителей, кроме тривиальных единиц, которые представляют собой круговые числа — делители 1. Нет проблемы и в разложении кругового числа или любого другого алгебраического числа на простые множители. Нужно просто делить число, пока не закончатся делители. Существует простой способ доказать, что эта процедура конечна, и когда она завершится, каждый делитель окажется простым. Так в чем же проблема? В единственности. Если вы повторите процедуру, выбирая по пути иные решения, вы вполне можете получить другой набор простых делителей.

На первый взгляд, трудно представить себе такую возможность. Простые делители — наименьшие возможные кусочки, на которые можно разбить число. Это как взять собранную из «Лего» игрушку и разобрать на кирпичики. Если бы существовал другой способ сделать это, то, в конце концов, оказалось бы, что мы разделили один из кирпичиков еще на несколько деталей. Но тогда кирпичик не был бы кирпичиком. К несчастью, аналогия с «Лего» обманчива. Алгебраические числа ведут себя не так. Они больше похожи на кирпичики с мобильными связями, способные соединяться между собой в разных сочетаниях. Разбейте кирпичик одним способом — получите одни составные части, которые сцепляются друг с другом и дальше уже не делятся. Разбейте его иначе — и получите еще один набор с теми же свойствами, но уже других деталей.

Я приведу два примера. В первом будут только обычные целые числа. Он несложен для понимания, но обладает некоторыми нерепрезентативными чертами. А затем я покажу вам настоящий пример.

Представьте, что мы живем во Вселенной, где существуют только числа 1, 5, 9, 13, 17, 21, 25 и т. д. — числа, которые в нашей нынешней Вселенной имели бы вид $4k + 1$. Если перемножить два таких числа, получится еще одно число такого же вида. Определим такое число как «простое», если оно не является произведением двух меньших чисел *того же вида*. К примеру, число 25 — не простое, поскольку равняется 5×5, а 5 тоже есть в нашем списке. Но число 21 — простое в этом новом смысле, потому что его обычных делителей (3 и 7) в списке нет. Они имеют вид $4k + 3$, а не $4k + 1$. Несложно убедиться, что любое число заданного вида есть произведение простых (в новом смысле) чисел. Причина в том, что множители, если они существуют, должны становиться меньше, и со временем процесс факторизации непременно остановится. Когда это произойдет, полученные множители будут простыми.

Однако такое разложение на простые множители не единственно. Рассмотрим число 4389. $4389 = 4 \times 1097 + 1$, т. е. это число интересующего нас вида. Вот три различных разложения на множители заданного вида:

$$4389 = 21 \times 209 = 33 \times 133 = 57 \times 77.$$

Я утверждаю, что, согласно принятому нами определению, все эти множители простые. К примеру, 57 — простое число, так как его обычные делители 3 и 19 не относятся к требуемому виду. То же можно сказать о числах 21, 33, 77, 133 и 209. Теперь мы можем объяснить неединственность разложения на простые множители. В обычных целых числах

$$4389 = 3 \times 7 \times 11 \times 19,$$

и все эти числа «не того» вида, они нам не подходят и имеют вид $4k + 3$. Три различных разложения на простые в этом новом смысле числа возникают при трех разных вариантах группировки этих чисел в пары:

$$(3 \times 7) \times (11 \times 19), (3 \times 11) \times (7 \times 19), (3 \times 19) \times (7 \times 11).$$

Мы вынуждены брать эти числа парами, так как два числа вида $4k + 3$ при перемножении дают число вида $4k + 1$.

Этот пример показывает, что аргумент «множители должны быть единственными, поскольку они минимальны», в данном случае не работает. Правда, здесь есть числа и меньше ($21 = 3 \times 7$, к примеру), но эти числа не принадлежат к интересующей нас системе. Главная же причина того, что этот пример не является полностью репрезентативным, заключается в том, что, хотя при умножении числа вида $4k + 1$ дают числа того же вида, при сложении этого не происходит. К примеру, $5 + 5 = 10$, но 10 — не число нужного нам вида. Поэтому, говоря языком абстрактной алгебры, мы имеем дело не с кольцом.

Второй пример не имеет этого недостатка, но он более сложен. Это кольцо алгебраических целых для многочлена $x^2 - 15$. В это кольцо входят все числа $a + b \quad 15$, где a и b целые. В нем число 10 имеет два варианта разложения:

$$10 = 2 \times 5 = (5 + \quad 15) \times (5 - \quad 15).$$

Можно доказать, что все четыре множителя ($2, 5, 5 + \quad 15$, $5 - \quad 15$) являются простыми[26].

Сегодня все это выглядит гораздо понятнее, чем в 1847 г., но математикам не потребовалось много времени, чтобы показать обоснованность сомнений Лиувилля. Через две недели после доклада Ванцель проинформировал Академию, что для небольших p единственность разложения соблюдается, но для 23-й степени его метод доказательства уже не годится. Вскоре после этого Лиувилль доложил Академии, что единственность разложения на простые множители *не соблюдается* для круговых целых чисел, соответствующих $p = 23$. (Эрнст Куммер открыл этот факт тремя годами раньше, но никому не сказал, поскольку искал способ обойти это препятствие.)

Доказательство Ламе работало для небольших значений p, включая некоторые новые (11, 13, 17, 19), но в общем случае неизбежно рассыпалось. Это был наглядный урок: нельзя принимать правдоподобные математические утверждения на веру, даже если они кажутся очевидными. Может оказаться, что они вообще неверны.

Куммер тоже искал доказательство Великой теоремы Ферма, и мысль его работала примерно в том же направлении, что и у Ламе. Он вовремя заметил потенциальное препятствие и отнесся к нему серьезно: проверил и обнаружил, что оно губит этот подход к доказательству. Он нашел конкретный пример неединственного разложения на простые делители для круговых чисел на основе корней 23-й степени из единицы. Но Куммер был не из тех, кто легко сдается, и ему удалось обойти препятствие или по крайней мере смягчить худшие его следствия. Его идею можно продемонстрировать особенно наглядно на примере все тех же чисел вида $4k + 1$. Чтобы сделать разложение по-прежнему единственным, достаточно добавить кое-какие *новые* числа, не принадлежащие к интересующей нас системе. Для этого примера нам нужны недостающие числа вида $4k + 3$. А можно не мелочиться и добавить к тому же четные целые числа; тогда мы получим множество целых чисел, замкнутое относительно сложения и умножения.

Иными словами, при сложении или умножении двух целых чисел результат тоже будет целым.

Куммер предложил другой вариант этой же идеи. К примеру, чтобы восстановить единственность разложения на простые множители в кольце чисел $a + b\ \ 15$, достаточно добавить к нему еще одно число, а именно $\ \ 5$. Далее выясняется, что, чтобы получить кольцо, мы должны добавить еще $\ \ 3$. Теперь

$$2 = (\ \ 5 + \ \ 3) \times (\ \ 5 - \ \ 3), 5 = \ \ 5 \times \ \ 5$$

и

$$5 + \ \ 15 = \ \ 5 \times (\ \ 5 + \ \ 3), 5 - \ \ 15 = \ \ 5 \times (\ \ 5 - \ \ 3).$$

Таким образом, при разных вариантах группировки четырех чисел $\ \ 5$, $\ \ 5$, $\ \ 5 + \ \ 3$, $\ \ 5 - \ \ 3$ возникает два варианта факторизации.

Куммер назвал эти новые множители идеальными числами, поскольку в его общих формулировках они вообще не считались числами в полной мере. Они были символами, которые вели себя в значительной степени как числа. Он доказал, что любое круговое целое число может быть единственным образом разложено на простые идеальные числа. Довольно тонкая схема: ни круговые числа, ни идеальные числа сами по себе не имели единственного разложения на простые множители. Но если воспользоваться идеальными числами как ингредиентами разложения для круговых чисел, то результат получался единственно возможным.

Позже Рихард Дедекинд нашел более цивилизованную интерпретацию процедуры Куммера, и ею мы пользуемся до сих пор. Каждому идеальному числу вне интересующего нас кольца он поставил в соответствие некий *набор* чисел внутри кольца. Этот набор он назвал идеалом. Каждое число в кольце определяет идеал: в него входят все числа, кратные данному. Если разложение на простые множители единственно, таков и каждый идеал. Если разложение не единственно, то возникают дополнительные идеалы. Мы можем определить произведение и сумму идеалов, а также простые идеалы, и Дедекинд доказал, что разложение *идеалов* на простые множители единственно для всех колец алгебраических целых чисел. Это позволяет предположить, что при решении большинства задач работать следует с идеалами, а не с самими алгебраическими числами. Конечно, здесь не обходится без новых сложностей, но альтернативой такому методу, как правило, является тупик.

Куммер научился работать со своими идеальными числами — по крайней мере научился достаточно хорошо, чтобы доказать вариант Великой теоремы Ферма при некоторых дополнительных предположениях. Но остальным смертным идеальные числа показались сложными и даже слегка загадочными. Однако, если посмотреть с позиции Дедекинда, идеальные числа разумны и полезны, и алгебраическая теория чисел начала свой путь. Из нее, в частности, возникла весьма важная идея о том, как можно измерить степень неединственности разложения в кольце алгебраических целых чисел. Каждому такому кольцу ставится в соответствие целое число, именуемое *классом*. Если класс равен 1, разложение на простые множители однозначно; в противном случае — нет. Чем выше класс, тем «менее однозначно» (почти в буквальном смысле) разложение на простые множители.

Возможность количественно оценить неоднозначность разложения стала серьезным шагом вперед: при помощи некоторых дополнительных усилий она спасала стратегию Ламе, но лишь в некоторых случаях. В 1850 г. Куммер объявил, что может доказать Великую теорему Ферма для большого числа простых чисел, которые он назвал регулярными. Из всех простых чисел до 100 к ним не относятся только 37, 59 и 67. Для всех остальных простых чисел до этого предела и очень многих после Куммер доказал Великую теорему Ферма. Для определения регулярного простого числа необходимо привлечь понятие класса: простое число регулярно, если оно не является делителем номера класса соответствующего кольца круговых целых чисел. Так что для регулярного простого числа разложение хотя и не является единственным, но это не затрагивает существенным образом интересующее нас простое число.

Куммер утверждал, что существует бесконечно много регулярных простых чисел, но это утверждение до сих пор остается недоказанным. По иронии судьбы, в 1915 г. К. Йенсен доказал, что существует бесконечно много иррегулярных простых чисел. Неожиданный критерий регулярности простых чисел выявился в связи с математическим анализом. В нем задействована последовательность чисел, открытая независимо японским математиком Сэки Такакадзу и швейцарским математиком Якобом Бернулли и известная как числа Бернулли. Этот критерий показывает, что в первый десяток иррегулярных простых чисел входят 37, 59, 67, 101, 103, 131, 149, 157, 233 и 257. Углубившись в структуру круговых чисел, Дмитрий Мириманов разобрался с первым из иррегулярных простых чисел — 37 — в 1893 г. К 1905 г. он доказал теорему Ферма до степеней p вплоть до 257. Гарри Вандивер разработал компьютерный алгоритм, который позволил расширить этот предел. При помощи этих методов Джон Селфридж и Бари Поллак в 1967 г. доказали теорему вплоть до 25 000-й степени, а С. Вагстафф в 1976 г. повысил этот предел до 100 000.

Свидетельства истинности Великой теоремы Ферма накапливались и накапливались, но главным, пожалуй, было то, что в случае ее ложности контрпример, т. е. пример, наглядно демонстрирующий нарушение теоремы, оказался бы настолько сложным, что никто и никогда не сумел бы его отыскать. Еще из результатов работы ученых следовало, что методы, такие как у Куммера, сталкивались с теми же проблемами, что и работы более ранних исследователей: большие степени требовали особых очень сложных процедур и особого обращения. Так что эта линия атаки постепенно затормозилась и сошла на нет.

Если при решении математической задачи вы оказались в тупике, последуйте совету Пуанкаре: отвлекитесь и займитесь чем-нибудь другим. Если вам повезет и ветер будет попутным, у вас рано или поздно появится новая идея. Специалисты по теории чисел вряд ли осознанно следовали этому совету, но тем не менее они поступали именно так. Как и утверждал Пуанкаре, такая тактика срабатывала. Некоторые специалисты по теории чисел перенесли внимание на эллиптические кривые (см. главу 6). По иронии судьбы, со временем именно в этой области математики выявились поразительные и неожиданные связи с Великой теоремой Ферма, которые и привели, в конце концов, к доказательству Уайлса. Для описания этих связей нам потребуется еще одно понятие — модулярной функции. С этого момента наше обсуждение приобретет несколько технический характер, но за всем этим стоит вполне разумная история, да нам и нужны-то лишь самые общие положения. Следите за моими рассуждениями.

В главе 6 мы видели, что теория эллиптических функций сильно повлияла на развитие комплексного анализа. В 1830-е гг. Жозеф Лиувилль открыл, что разновидностей эллиптических функций не так уж много. Для любых двух периодов существует особая эллиптическая функция, известная как функция Вейерштрасса, и любая другая эллиптическая функция с теми же двумя периодами является просто ее вариантом. Тем самым подразумевается, что из функций с двойной периодичностью достаточно разобраться в функциях Вейерштрасса — по одной на каждую пару периодов.

Геометрически двойную периодическую структуру эллиптической функции можно интерпретировать как решетку на комплексной плоскости: это все комбинации вида $mu + nv$ двух периодов u и v (см. рис. 30). Если мы возьмем комплексное число z и добавим к нему одну из точек нашей решетки, то значение эллиптической функции в новой точке будет тем же, что и в первоначальной. Иными словами, эллиптическая функция обладает той же симметрией, что и описанная решетка.

Аналитики открыли гораздо более богатый источник симметрий комплексной плоскости, известный как преобразования Мёбиуса. Эти преобразования превращают z в $(az + b)/(cz + d)$, где a, b, c, d — комплексные константы. Симметрии, определяемые решеткой, представляют собой особые случаи преобразований Мёбиуса, но существуют и другие. Однако в более общем случае тоже присутствует набор точек, аналогичный рассмотренной нами решетке. Решетка определяет на евклидовой плоскости ячеистую структуру: достаточно взять в виде ячейки параллелограмм и поместить его углы в узлы решетки (см. рис. 26 и 30). При помощи преобразований Мёбиуса мы можем построить ячеистую структуру в подходящей неевклидовой геометрии, на гиперболической поверхности. Мы можем установить тождественность этой поверхности и части комплексной плоскости, где прямые заменяются дугами окружностей.

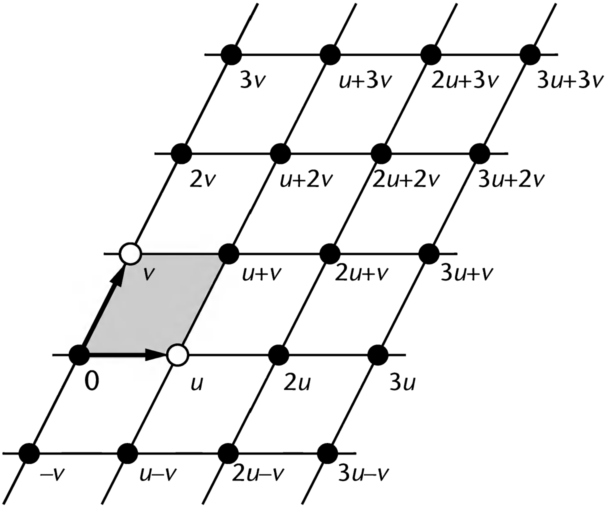


Рис. 30. Решетка образуется из всех целых комбинаций двух периодов

В гиперболической геометрии существуют весьма симметричные ячеистые структуры. Для каждой из них можно построить комплексные функции, которые на каждой ячейке повторяют свои значения. Такие функции известны как модулярные и представляют собой естественное обобщение эллиптических функций. Гиперболическая геометрия — очень насыщенная область математики, и диапазон ячеистых структур здесь намного шире, чем на евклидовой плоскости. Поэтому специалисты по комплексному анализу всерьез заинтересовались неевклидовой геометрией. При этом выявилась глубокая связь между математическим анализом и теорией чисел. Модулярные функции играют для эллиптических кривых ту же роль, что тригонометрические функции для окружности.

Напомню, что единичная окружность состоит из точек (x, y), таких, что $x^2 + y^2 = 1$. Пусть A — действительное число, и

$$x = \cos A, \ y = \sin A.$$

Тогда определение синуса и косинуса говорит о том, что данная точка лежит на единичной окружности. Более того, любая точка единичной окружности имеет такую форму. Говоря математическим языком, эти тригонометрические функции представляют окружность в параметрическом виде. Что-то очень похожее происходит и с модулярными функциями. Если мы определим x и y при помощи подходящих модулярных функций параметра A, то соответствующая точка будет лежать на эллиптической кривой — одной и той же эллиптической кривой, какое бы значение ни принимал параметр A. Существуют и более абстрактные способы сформулировать вышеизложенное, и специалисты пользуются именно ими, потому что они удобнее, но этот вариант позволяет выявить аналогию с тригонометрическими функциями и окружностью. Эта связь порождает свою эллиптическую кривую для каждой модулярной функции, а разнообразие модулярных функций громадно — ведь это все симметричные ячеистые структуры на гиперболической поверхности. Итак, огромное количество эллиптических кривых может быть соотнесено с модулярными функциями. Но какие эллиптические кривые можно получить таким способом? Именно этот вопрос оказался главным.

Впервые это «недостающее звено» привлекло внимание ученых в 1975 г., когда Ив Эллегуар обратил внимание на занятную связь между Великой теоремой Ферма и эллиптическими кривыми. Герхард Фрей в двух статьях, опубликованных в 1982 и 1986 гг., развил эту идею. Пусть p, как всегда, нечетное простое число. Предположим — в надежде прийти к противоречию, — что существуют ненулевые целые числа a, b и c, удовлетворяющие уравнению Ферма, так что $a^p + b^p = c^p$. А теперь с надлежащей помпой извлечем из шляпы заранее припасенного кролика: рассмотрим эллиптическую кривую

$$y^2 = x\,(x - a^p)\,(x - b^p).$$

Эта кривая называется эллиптической кривой Фрея. Фрей применил к ней механизм работы с эллиптическими кривыми и получил цепочку еще более причудливых совпадений. Его гипотетическая эллиптическая кривая выглядит и правда очень странно. На первый взгляд, она вообще лишена смысла. Фрей доказал, что смысла в ней настолько мало, что она не может существовать. И это обеспечивает нам желанное противоречие и тем самым, разумеется, доказывает Великую теорему Ферма.

Однако в этом доказательстве есть пробел, и Фрей прекрасно знал о нем. Чтобы доказать, что такая эллиптическая кривая не существует, необходимо показать, что если бы она существовала, то была бы модулярной, т. е. одной из тех кривых, что возникают из модулярных функций. Мы только что убедились, что таких кривых множество; на тот момент никому не удавалось отыскать хотя бы одну эллиптическую кривую, которая не была бы модулярной. Казалось логичным, что и кривая Фрея должна быть модулярной, но это была гипотетическая кривая, коэффициенты *a*, *b* и *c* не были известны. К тому же, если бы кривая и правда была модулярной, то она просто не могла бы существовать. Был, однако, один способ раз и навсегда разобраться со всеми этими вопросами: доказать, что *все* эллиптические кривые модулярны. Тогда кривая Фрея, гипотетическая или нет, тоже была бы модулярной, если бы существовала. А если бы ее не было, то доказательство от этого никак бы не пострадало.

Утверждение, что всякая эллиптическая кривая является модулярной, называется гипотезой Таниямы–Симуры. Она названа в честь двух японских математиков Ютаки Таниямы и Горо Симуры. Встретились они случайно: оба одновременно с одной и той же целью хотели получить в университетской библиотеке одну и ту же книгу. Результатом же стало долгое сотрудничество. В 1955 г. Танияма был в Токио на математической конференции, где молодым участникам предложили составить список открытых вопросов. Танияма предложил четыре вопроса, и все они были связаны с отношениями между модулярными функциями и эллиптическими кривыми. Еще до этого он вычислил некоторые числа, связанные с конкретной модулярной функцией, и заметил, что в точности те же числа появлялись в связи с конкретной эллиптической кривой. Подобные совпадения часто свидетельствуют о том, что все это вовсе не совпадение и что замеченным фактам должно быть какое-то разумное объяснение. Сегодня мы знаем: равенство этих чисел напрямую означает, что эллиптическая кривая является модулярной, более того, именно так чаще всего определяется модулярность в специальной литературе. Так или иначе, Танияма был достаточно заинтригован, чтобы рассчитать соответствующие числа еще для нескольких модулярных функций и выяснить, что они тоже соответствуют конкретным эллиптическим кривым.

Он заинтересовался, не найдется ли подобной черты у каждой эллиптической кривой. Специалисты в этой области в большинстве своем считали, что это слишком хорошо, чтобы быть правдой, — бесплодная мечта, в пользу которой нет почти никаких свидетельств. Симура был одним из немногих, кто считал, что эта гипотеза достойна серьезного рассмотрения. Но в 1957–1958 гг. Симура уехал на год в Принстон, а Танияма, пока его не было, покончил с собой. В оставленной им записке, в частности, говорилось: «*Причину моего самоубийства я не могу и сам понять, но это не результат какого-то конкретного события, нет никаких особенных причин. Единственное, что я точно знаю, — я потерял веру в будущее*».

Примерно месяц спустя его невеста Мисако Судзуки тоже покончила с собой. В ее прощальной записке было сказано: «Теперь, когда его нет, я тоже должна уйти, чтобы присоединиться к нему».

Симура продолжил работу над гипотезой. По мере того как накапливались свидетельства в ее пользу, он начал склоняться к мысли о том, что она действительно может оказаться верной. Большинство других специалистов были с ним не согласны. Саймон Сингх рассказывает об интервью с Симурой, в котором тот вспоминал, как пытался объяснить все это одному из коллег:

> «Профессор поинтересовался: “Я слышал, вы предполагаете, что некоторые эллиптические уравнения могут быть связаны с модулярными формами”.
> “Нет, вы не понимаете, — ответил Симура. — Речь не о *некоторых* эллиптических уравнениях, так ведут себя *все* эллиптические уравнения!”»

Несмотря на общий скептицизм, Симура продолжал упорно работать, и с годами это предположение приобрело достаточную респектабельность, чтобы о нем стали говорить как о гипотезе Таниямы–Симуры. Затем Андре Вейль, один из крупнейших специалистов по теории чисел XX столетия, нашел дополнительные свидетельства в ее пользу, опубликовал их и высказал уверенность в том, что она на самом деле вполне может быть верна. После этого гипотезу стали называть гипотезой Таниямы–Симуры–Вейля. Вообще, название ее окончательно не устоялось, и в публикациях о ней можно встретить самые разные комбинации имен трех этих математиков. Я буду придерживаться названия «гипотеза Таниямы–Симуры».

В 1960-е гг. еще один математический тяжеловес Роберт Ленглендс понял, что гипотезу Таниямы–Симуры можно рассматривать как один из элементов куда более обширной и амбициозной программы, способной объединить алгебраическую и аналитическую теорию чисел. Он сформулировал целый набор гипотез, связанных с этой идеей и известных сегодня как программа Ленглендса. Она была еще более спекулятивна, чем гипотеза Таниямы–Симуры, но обладала неотразимой элегантностью: подобная математика настолько красива, что просто обязана быть истинной. В течение последующего десятилетия математический мир постепенно оценил красоту программы Ленглендса, и ее исполнение начали воспринимать как одну из главных целей алгебраической теории чисел. Программа Ленглендса представляется верным направлением развития, если, конечно, кому-то удастся сделать в этом направлении первый шаг.

В 1980-е Фрей заметил, что применение гипотезы Таниямы–Симуры к его эллиптической кривой означало бы доказательство Великой теоремы Ферма. Однако к тому времени выявилась еще одна проблема с его идеей. Когда в 1984 г. он прочел лекцию на эту тему, аудитория заметила прореху в ключевом аргументе: его кривая настолько необычна, что просто не может быть модулярной. Один из ведущих специалистов в этой области Жан-Пьер Серр быстро закрыл прореху, но для этого ему пришлось задействовать еще один результат, также нуждавшийся в доказательстве, — специальную гипотезу о понижении уровня. К 1986 г., однако, Кен Рибет доказал эту гипотезу. Теперь единственным препятствием на пути к доказательству теоремы Ферма была гипотеза Таниямы–Симуры, и мнение математического сообщества начало потихоньку смещаться. Серр предсказал, что Великая теорема Ферма, вероятно, будет доказана в течение ближайших десяти лет или около того. Как именно доказана, оставалось вопросом, но в воздухе уже витала общая уверенность в успехе: методики, связанные с модулярными функциями, обретали такую мощь, что очень скоро кто-нибудь должен был реализовать наконец подход Фрея.

Этим кем-то оказался Эндрю Уайлс. В телепрограмме, целиком посвященной доказательству теоремы Ферма, он рассказал:

> «Мне было 10 лет, когда я нашел книгу по математике, в которой рассказывалось немного об истории этой задачи [Великой теоремы Ферма], — что один человек написал ее 300 лет назад, но никто никогда не видел ее доказательства, никто не знал, существует ли оно, и с тех пор люди искали его. Передо мной была задача, которую я, десятилетний мальчик, был в состоянии понять, но которую никто из великих математиков прошлого не смог решить. И с того момента я, конечно, пытался решить ее сам. Это был такой вызов, такая красивая задача».

В 1971 г. Уайлс получил в Оксфорде диплом по математике и переехал в Кембридж работать над докторской диссертацией. Его руководитель Джон Коутс сказал ему (и был совершенно прав), что теорема Ферма слишком сложна для докторской диссертации и отсоветовал за нее браться. Так что Уайлс занялся эллиптическими кривыми, которые тогда считались куда более многообещающим полем для исследований. К 1985 г. он был уже в Париже в Институте высших научных исследований — одном из ведущих мировых центров математических исследований. Большинство лучших ученых в тот или иной момент проходят через это учреждение, и, если вы математик, это великолепное место для работы и общения. В то время в Институте бывал и Рибет, и его доказательство специальной теоремы о понижении уровня буквально наэлектризовало

Уайлса. Теперь он мог заниматься в высшей степени респектабельным исследованием эллиптических кривых и пытаться доказать гипотезу Таниямы–Симуры — и в то же время стремиться к исполнению своей детской мечты доказать Великую теорему Ферма.

Поскольку все теперь знали о связи между этими областями, это вызывало некоторое беспокойство. Предположим, Уайлс сумел бы собрать почти полное доказательство, в котором оставалось бы лишь несколько небольших пробелов, требующих дополнительных усилий. Предположим также, что кто-то другой узнал бы об этом и заполнил оставшиеся пробелы. Тогда технически именно этот человек стал бы автором доказательства Великой теоремы Ферма. Как правило, математики так себя не ведут, но приз был слишком велик, и Уайлс благоразумно решил принять меры предосторожности. Он вел свои исследования в тайне, что математикам несвойственно. И дело было не в том, что Уайлс не доверял коллегам. Он просто не хотел допустить даже малейшей вероятности того, что кто-нибудь обойдет его перед финишной чертой.

Семь лет Уайлс работал на чердаке своего дома, где был оборудован кабинет. Только жена и непосредственный начальник знали, чем именно он занимается. В тишине и уединении он атаковал задачу всеми методами, какие мог вспомнить и освоить, и, в конце концов, стены крепости начали сотрясаться под его ударами. В 1991 г. Коутс познакомил его с новыми результатами и доказательствами, полученными Маттеусом Флахом. Осада продвигалась успешно: по крепостной стене уже пошли трещины.

К 1993 г. работа над доказательством была завершена. Оставалось лишь представить его миру. Однако Уайлс все еще осторожничал: ему не хотелось рисковать и объявлять о своем достижении только для того, чтобы тут же выявилась какая-нибудь ошибка. Примерно так произошло с Йоити Мияокой в 1988 г.: средства массовой информации поспешили разнести по всему миру его заявление о том, что получено доказательство Великой теоремы Ферма, в котором очень быстро была обнаружена ошибка. Поэтому Уайлс решил провести серию из трех лекций в кембриджском Институте Исаака Ньютона — недавно организованном международном исследовательском центре по математике. Тема лекций звучала безобидно: «Модулярные формы, эллиптические кривые и теория Галуа». Иносказание, однако, мало кого обмануло: все понимали, что Уайлс собирается объявить о серьезном открытии.

На третьей лекции Уайлс коротко изложил доказательство одного особого случая гипотезы Таниямы–Симуры. Он выяснил, что можно обойтись и чуть менее строгим утверждением. Достаточно доказать, что кривая Фрея, если она существует, должна принадлежать к особому классу эллиптических кривых, известных как «полустабильные», а затем доказать, что все кривые *этого класса* модулярны. Уайлс доказал и то и другое. В конце той лекции он записал на доске следствие — дополнительную теорему, которая непосредственно следует из того, что было только что доказано. Этим следствием была Великая теорема Ферма.

Симура, услыхав о заявлении Уайлса, высказался кратко и по существу: «Я же говорил!»

Но если бы все было так однозначно! У судьбы в запасе нашелся еще один неожиданный поворот. Доказательство нуждалось в одобрении и признании специалистов, и в процессе его рассмотрения, как обычно, выявилось несколько моментов, по которым требовались дополнительные разъяснения. Уайлс справился с большей частью подобных комментариев, но один из них заставил его задуматься. В конце 1993 г. он опубликовал заявление: отозвал свои претензии на доказательство Великой теоремы до тех пор, пока ему не удастся заполнить выявленный логический пробел. Но теперь работать ему приходилось в обстановке полной публичности, т. е. произошло именно то, чего он надеялся избежать.

К марту 1994 г. исправленное доказательство не появилось, и Фальтингс выразил широко распространенное в математическом сообществе мнение: «Если бы [исправить доказательство] было просто, он бы уже решил эту проблему. Строго говоря, то, что было заявлено, не доказательство». Вейль заметил: «Я считаю, что у него есть несколько хороших идей, но доказательства нет… Доказать Великую теорему Ферма — это как взобраться на Эверест. Если кто-то хочет покорить Эверест и не доходит до вершины 100 ярдов, это означает, что он не покорил Эверест». Каждый мог легко представить себе, чем все кончится. Все это уже было. Доказательство рухнуло, его придется полностью пересматривать, и теорема Ферма останется непобежденной до следующего сражения.

Но Уайлс отказывался признавать поражение. К поискам присоединился его бывший студент Ричард Тейлор. Корень проблемы теперь был ясен: результаты Флаха не слишком хорошо подходили к этой задаче. Уайлс и Тейлор попытались доработать методы Флаха, но ничего не получалось. Затем Уайлса осенило: он внезапно понял, в чем состоит главное препятствие. «Я понял, что та самая штука, из-за которой перестал работать метод Флаха, может заставить работать другой метод, который я тоже когда-то пробовал». Солдаты, осаждающие замок, вдруг поняли, что их таран никогда не проломит ворота, поскольку осажденные постоянно сбрасывают на него большие камни, но можно этими самыми камнями зарядить требушет и пробить ворота иначе.

К апрелю 1995 г. новое доказательство было завершено, и на этот раз в нем не оказалось ни прорех, ни ошибок. Оно было опубликовано в виде двух статей в *Annals of Mathematics* — одном из самых уважаемых математических журналов. Уайлс стал мировой знаменитостью, удостоился нескольких престижных премий и рыцарского звания… и вернулся к своим исследованиям. Его жизнь почти не изменилась.

По-настоящему важно в достижении Уайлса вовсе не доказательство Великой теоремы Ферма как таковое. Я уже говорил, что от того, верна эта теорема или нет, практически ничего не зависит. Если бы кто-то отыскал три 100-значных числа и 250-значный простой показатель степени, из которых сложился бы контрпример к утверждению Ферма, то ни одна важная область математики от этого не пострадала бы. Конечно, прямой компьютерный расчет не осилил бы таких больших чисел, так что вам пришлось бы проявить недюжинный ум, чтобы отыскать что-нибудь подобное, но отрицательный результат в данном случае не вызвал бы ни у кого особой тревоги.

Подлинное значение найденного Уайлсом решения проблемы заключается в доказательстве гипотезы Таниямы–Симуры для полустабильного случая. Не прошло и шести лет, как Кристоф Брейль, Брайан Конрад, Фред Даймонд и Ричард Тейлор расширили методы Уайлса и разобрались не только с полустабильными, но и со всеми остальными эллиптическими кривыми. Они доказали полный вариант гипотезы Таниямы–Симуры, и теория чисел уже никогда не будет прежней. Теперь, если кто-то столкнется с эллиптической кривой, она гарантированно будет модулярной, что открывает перед исследователями множество новых аналитических методов. Эти методы уже используются для решения других задач теории чисел, а в будущем их появится еще больше.

8

Орбитальный хаос

Задача трех тел

Если верить старой шутке, то о продвинутости физической теории можно судить по тому, с каким количеством взаимодействующих тел она не в состоянии разобраться. Закон всемирного тяготения Ньютона сталкивается с проблемами уже на трех телах. Общая теория относительности с трудом справляется с двумя. Квантовая теория и для одного-то тела непомерно сложна, а квантовая теория поля попадает в беду даже там, где тел нет вообще — в вакууме. В этой шутке, как и во многих других, есть доля истины[27]. Так, над задачей гравитационного взаимодействия всего лишь трех тел, которые вроде бы подчиняются ньютонову обратно-квадратичному закону тяготения, математический мир бился не одну сотню лет. И до сих пор бьется, если говорить о красивой формуле для орбит этих тел. Правда, сегодня мы знаем, что динамика трех тел хаотична — настолько нерегулярна, что несет в себе элементы случайности.

Все это выглядит достаточно странно на фоне поразительного успеха гравитационной теории Ньютона, которая объяснила, помимо всего прочего, движение планет вокруг Солнца. Ответом было то, что Кеплер уже вывел эмпирически из астрономических наблюдений Марса: эллипс. Здесь задействованы только два тела: Солнце и планета. Очевидный следующий шаг заключается в том, чтобы записать уравнение для орбит трех тел и решить его. Но у этих орбит нет точных геометрических характеристик, нет даже формулы в геометрических координатах. До конца XIX в. о движении трех небесных тел было известно очень немного, даже в том случае, если одно из них настолько мало, что его массой можно пренебречь.

С тех пор наши представления о динамике трех (или более) тел сильно обогатились, а понимание того, насколько сложен этот вопрос и почему, выросло. Это может показаться ретроградством, но иногда, чтобы продвинуться вперед, лучше всего организовать стратегическое отступление и попробовать другие методы. Для задачи трех тел этот план кампании неоднократно приносил успех в случаях, когда лобовая атака безнадежно завязла бы в обороне.

Древние люди не могли не замечать, что Луна постепенно сдвигается по ночному небу относительно звездного фона. Звезды тоже вроде бы движутся, но все вместе, как единое целое, как крохотные световые точки на громадном вращающемся куполе небес. Луна же, очевидно, совершенно особый объект: это великолепный сияющий диск, меняющий форму от узенького полумесяца новой Луны до полного круга и обратно. Это не светящаяся точка, как звезды.

Некоторые светящиеся точки тоже не подчиняются общим правилам. Они блуждают по небу. Они не меняют своего положения относительно звезд так быстро, как Луна, но все же не обязательно слишком долго наблюдать за небом, чтобы заметить, что они движутся отдельно. Пять таких «блуждающих звезд» видимы невооруженным глазом. Греки назвали их *планетами* (*planetes*) — блуждающими. И, конечно, это и есть планеты (*planets*), сегодня мы называем их Меркурием, Венерой, Марсом, Юпитером и Сатурном — в честь римских богов. С помощью телескопов мы узнали о существовании еще двух: Урана и Нептуна. Плюс наша Земля, разумеется. А вот Плутон уже не считается планетой благодаря спорному решению по терминологии, принятому в 2006 г. Международным астрономическим союзом.

Изучая небеса, древние философы, астрономы и математики пришли к выводу, что планеты блуждают по небу не беспорядочно. Они следуют собственными извилистыми, но достаточно предсказуемыми путями и через строго определенные промежутки времени возвращаются примерно в ту же позицию на ночном небе. Сегодня мы объясняем эти маршруты периодическим движением по замкнутой орбите плюс некоторым влиянием со стороны собственного орбитального движения Земли. Мы признаем также, что периодичность здесь не строгая, но близкая к тому. У Меркурия путь вокруг Солнца занимает около 88 суток, а у Юпитера — почти 12 лет. Чем дальше от Солнца находится планета, тем больше времени у нее уходит на полный оборот вокруг светила.

Первую количественно точную модель движения планет дала система Птолемея. Свое название она получила в честь Клавдия Птолемея, описавшего ее в своем трактате «Альмагест» (что означает «Величайшее построение») около 150 г. н. э. Это геоцентрическая, т. е. с Землей в центре мироздания, модель, в которой все небесные тела движутся вокруг нашей планеты так, будто поддерживаются серией гигантских сфер, каждая из которых вращается с постоянной скоростью вокруг неподвижной оси. Комбинации множества вращающихся сфер требовались для того, чтобы представить сложное движение планет в виде космического идеала равномерного движения по кругу — экватора сферы. Если сфер достаточно, а их оси и скорости выбраны правильно, эта модель очень точно отражает реальность.

Николай Коперник доработал схему Птолемея в нескольких отношениях. Самым радикальным изменением было то, что он заставил все тела, кроме Луны, обращаться не вокруг Земли, а вокруг Солнца, что сильно упростило модель. Это не понравилось католической церкви, но со временем научные взгляды взяли верх, и все образованные люди приняли как данность то, что Земля обращается вокруг Солнца. В 1596 г. Кеплер защищал систему Коперника в своей книге «Тайна мира» (Mysterium Cosmographicum), в которой описал связь между расстоянием от Солнца до планеты и ее орбитальным периодом. Если двигаться от Солнца наружу, прирост периода обращения вдвое превышает прирост расстояния от светила. Позже Кеплер решил, что это соотношение слишком неточно, чтобы быть верным, но именно оно посеяло семена будущих более точных выводов. Кроме того, Кеплер объяснил расстояния между планетами через пять правильных многогранников, аккуратно вписанных друг в друга и разделенных удерживающими их сферами. Пять многогранников поясняли, с его точки зрения, почему планет пять, но сегодня мы знаем о существовании восьми планет, так что данная особенность уже не является аргументом в пользу такой гипотезы. Вообще говоря, существует 120 способов последовательно вписать пять правильных многогранников друг в друга, и, вполне возможно, один из этих способов даст соотношение, близкое к соотношению орбит. Так что это просто случайное приближение, приписывающее природе искусственную и бессмысленную закономерность.

В 1600 г. астроном Тихо Браге нанял Кеплера в качестве ассистента, но их совместная работа продлилась недолго. После смерти Браге Кеплер получил место придворного математика при дворе императора Рудольфа II. В свободное время он анализировал результаты наблюдений Браге за Марсом. Одним из результатов этой работы стала «Новая астрономия» (Astronomia Nova), которая вышла в 1609 г. и представила миру два закона планетарного движения. Первый закон Кеплера гласит, что планеты двигаются по эллипсам — он установил этот факт для Марса, и казалось вероятным, что другие планеты подчиняются тому же закону. Первоначально он считал, что данные хорошо лягут на яйцевидную орбиту, но с этим ничего не получилось; тогда он попробовал эллипс. После проверки эллипс тоже был отвергнут, и Кеплер нашел другое математическое описание формы орбиты, однако в конце концов понял, что его описание — всего лишь иной способ определения эллипса.

> «Я отложил [новое определение] в сторону и вернулся к эллипсам, будучи уверенным, что это совершенно иная гипотеза, тогда как обе они, как я докажу в следующей главе, суть одно и то же… Ах, каким глупым я был!»

Второй закон Кеплера гласит, что радиус-вектор планеты заметает за равные промежутки времени равные площади. В 1619 г. в работе «Гармония мира» (Harmonices Mundi) Кеплер завершил изложение своих трех законов куда более точным соотношением, связывающим расстояния и периоды: куб расстояния (большой полуоси эллипса) пропорционален квадрату периода обращения.

Можно сказать, что этим завершилась подготовка сцены к появлению на ней Исаака Ньютона. В работе 1687 г. «Математические начала натуральной философии» (Philosophiae Naturalis Principia Mathematica) Ньютон доказал, что три закона Кеплера эквивалентны единственному закону тяготения: два тела притягиваются друг к другу с силой, пропорциональной их массам и обратно пропорциональной квадрату расстояния между ними. Закон Ньютона обладал громадным преимуществом: он был применим к любой системе тел, сколько бы их ни было. Но за это приходилось платить: закон описывал орбиты не как геометрические формы, а как решения дифференциального уравнения, в которое входили, в частности, ускорения планет. Совершенно непонятно, как из такого уравнения определить форму планетарных орбит или, скажем, положение планет в заданный момент времени. Откровенно говоря, не совсем ясно даже, как найти эти самые ускорения планет. Тем не менее *неявно* вся эта информация в уравнении содержалась. Проблема заключалась в том, чтобы получить ее в явном виде. Кеплер уже проделал такую операцию для двух тел, и ответом стала эллиптическая орбита и скорость, при которой радиус-вектор каждой планеты описывает равные площади за равные промежутки времени.

Как же обстоит дело с тремя телами?

Хороший вопрос. Согласно закону Ньютона, все тела Солнечной системы притягивают друг друга. Более того, все тела во Вселенной притягивают друг друга. Но никто в здравом уме не стал бы пытаться записывать дифференциальные уравнения для каждого тела во Вселенной. Как всегда, чтобы продвинуться вперед, нужно было упростить задачу, но не слишком сильно. Звезды так далеки от нас, что их гравитационным влиянием на Солнечную систему можно пренебречь, если только вы не собираетесь описывать движение Солнца в Галактике или вращение самой Галактики. Движением Луны в значительной мере управляют два тела — Земля и Солнце — плюс некоторые тонкие эффекты от влияния других планет. В начале XVIII в. этот вопрос вышел за рамки чистой астрономии и приобрел практическое значение: ученые осознали, что движение Луны по небу можно использовать для навигации. (В те времена не было не только системы GPS, но и хронометров для определения долготы.) Но этот метод требовал более точных предсказаний, чем те, что позволяла сделать существующая теория. Очевидно, для начала следовало записать следствия из закона Ньютона для трех тел, которые в данном случае можно было рассматривать как точечные массы, поскольку планеты чрезвычайно малы по сравнению с расстояниями между ними. Затем следовало решить полученные дифференциальные уравнения. Однако методы, позволившие в задаче для двух тел перейти к эллипсам, в задаче для трех тел оказались неприменимы: добавление третьего тела портило всю картину. Несколько предварительных шагов сделать удалось, но затем вычисления зашли в тупик. В 1747 г. Жан д'Аламбер и Алексис Клеро, вечные соперники, приняли участие в конкурсе, объявленном Парижской академией наук по «задаче трех тел», которую оба пытались решить при помощи численных приближений. Задача для трех тел обрела название и вскоре стала одной из великих загадок математики.

Некоторые частные случаи этой задачи удавалось решить. В 1767 г. Эйлер обнаружил решения, в которых все три тела лежат на вращающейся прямой. В 1772 г. Лагранж нашел аналогичные решения для случая, когда тела образуют вращающийся равносторонний треугольник, который может расширяться или сжиматься. Оба решения оказались периодическими: тела повторяли одну и ту же последовательность движений до бесконечности. Однако даже кардинальное упрощение не позволяло получить хоть что-нибудь более общее. Можно было считать, что масса одного из тел пренебрежимо мала или что другие два тела движутся вокруг общего центра масс по идеальным окружностям (версия, известная как «ограниченная задача трех тел»), но найти точное решение уравнений все равно не удавалось.

В 1860 и 1867 гг. астроном и математик Шарль-Эжен Делоне пытался решить задачу для конкретного случая — системы Солнце–Земля–Луна — с использованием теории возмущений. Эта теория рассматривает действие солнечного притяжения на Луну как небольшие добавки, которые накладываются на действие земного притяжения. Делоне вывел приближенные формулы в виде сумм бесконечных рядов: результата сложения множества последовательных членов. Он опубликовал свои результаты в виде двух томов по 900 страниц в каждом. Эти тома были заполнены преимущественно формулами. В конце 1970-х гг. его расчеты были проверены с использованием компьютерной алгебры и подтвердились почти полностью: в них обнаружились всего две незначительные ошибки.

Это был поистине героический расчет, но ряд у Делоне сходился к своему пределу слишком медленно, чтобы этими выкладками можно было пользоваться на практике. Однако работа Делоне подтолкнула других математиков к поиску рядов, которые сходились бы быстрее. Она также вскрыла серьезное техническое препятствие, с которым неизменно встречается подобный подход: это препятствие — малые знаменатели. Некоторые члены последовательности представляют собой дроби, и знаменатель этих дробей вблизи резонанса (состояния, в котором периоды тел кратны друг другу) становится очень маленьким. К примеру, у трех внутренних спутников Юпитера — Ио, Европы и Ганимеда — периоды обращения вокруг планеты составляют 1,77, 3,55 и 7,15 суток, т. е. относятся один к другому почти точно как 1:2:4. Особенно мешает вычислениям секулярный резонанс, при котором кратны друг другу скорости поворота осей двух почти эллиптических орбит, — здесь при вычислении дроби с малым знаменателем погрешность становится очень большой.

Если задача трех тел сложна, то задача *n* тел, т. е. произвольного числа точечных масс, движущихся под действием ньютонового тяготения, безусловно, еще сложнее. Тем не менее природа представляет нам наглядный и очень важный пример: Солнечную систему. В нее входят восемь планет, несколько карликовых планет, таких как Плутон, и тысячи астероидов, в том числе довольно крупных. Это не говоря о спутниках планет, некоторые из которых — Титан, к примеру, — превосходят по размеру планету Меркурий. Таким образом, Солнечная система — это задача 10, или 20, или 1000 тел в зависимости от степени детализации.

Для краткосрочных прогнозов вполне достаточно численных аппроксимаций (в астрономии 1000 лет — это немного), а вот понять, как будет развиваться Солнечная система в ближайшие несколько сотен миллионов лет, — совсем другое дело. Но есть один серьезный вопрос, ответ на который зависит от подобных долгосрочных прогнозов: речь идет о стабильности Солнечной системы. Планеты в ней, судя по всему, обращаются по относительно стабильным, почти эллиптическим орбитам. Эти орбиты слегка изменяются, когда их возмущают другие планеты, так что период обращения и размеры эллипса могут чуть-чуть меняться. Можем ли мы быть уверены, что и в будущем не будет происходить ничего, кроме этого мягкого влияния? И так ли вела себя Солнечная система в прошлом, особенно на ранних стадиях развития? Останется ли она стабильной или какие-нибудь две ее планеты могут когда-нибудь столкнуться? Наконец, может ли планета оказаться выброшенной из системы прочь, на просторы Вселенной?

В 1889 г. королю Норвегии и Швеции Оскару II должно было исполниться 60 лет. Норвежский математик Геста Миттаг-Лефлер убедил короля объявить к юбилею конкурс на решение задачи *n* тел с немаленьким призом. Решение должно было представлять собой не точную формулу — к тому моменту было уже ясно, что это означало бы требовать слишком многого, — а некий сходящийся ряд. Пуанкаре, заинтересовавшийся конкурсом, решил начать с очень простой версии: ограниченной задачи трех тел, где масса одного из тел пренебрежимо мала, как, скажем, у пылинки. Если вы наивно примените закон Ньютона к такой пылинке, приложенная к ней сила будет равняться произведению масс, деленному на квадрат расстояния. При нулевой массе результат тоже будет равняться нулю. Это не слишком помогает, поскольку получается, что пылинка мирно летит своей дорогой, не взаимодействуя с остальными двумя телами. Вместо этого можно применить модель, в которой пылинка испытывает влияние остальных двух тел, а вот они полностью ее игнорируют. В этом случае орбиты двух массивных тел оказываются круговыми, и движутся они с постоянной скоростью. Вся сложность движения в такой системе приходится на пылинку.

Пуанкаре не решил задачу, поставленную королем Оскаром, — она была попросту слишком сложной. Но его методы были настолько новаторскими и продвинуться ему удалось так далеко, что приз он все же получил. Исследование было опубликовано в 1890 г. Из него явствовало, что даже ограниченная задача трех тел может не иметь предполагаемого решения. Пуанкаре разделил свой анализ на несколько отдельных случаев в зависимости от общих параметров движения. В большинстве случаев решение в виде ряда вполне можно было получить. Но был один случай, в котором орбита пылинки становилась чрезвычайно путанной.

Пуанкаре вывел эту неизбежную путаность при помощи некоторых других методов, над которыми работал в то время. Эти методы давали возможность описать решения дифференциальных уравнений, не решая их. Его «качественная теория дифференциальных уравнений» стала зерном, из которого выросла современная нелинейная динамика. Основной идеей, которая легла в основу новой теории, было исследование геометрии решений, точнее, их топологии — темы, глубоко интересовавшей Пуанкаре (см. главу 10). В такой интерпретации положения и скорости тел представляют собой координаты в многомерном пространстве. По мере того как идет время, первоначальное состояние системы движется в этом пространстве по некоей криволинейной траектории. Топология этого пути или даже системы всех возможных путей могут рассказать нам много полезного о решениях.

Периодическое решение, к примеру, представляет собой замкнутую траекторию в форме петли. По ходу времени состояние системы вновь и вновь проходит по этой траектории, бесконечно повторяя одно и то же поведение. Тогда и система является периодической. Пуанкаре предположил, что для удобного поиска подобных петель удобно было бы провести многомерную поверхность так, чтобы она рассекла петлю. Мы сегодня называем такую поверхность сечением Пуанкаре. Решения, берущие начало на этой поверхности, могут со временем вернуться на нее. Сама петля при этом возвращается в точности в ту же точку, а решения, проходящие через ближайшие к этой точки, всегда возвращаются на наше сечение примерно через один период. Так что периодическое решение можно интерпретировать как неподвижную точку на «отображении первого возвращения». Это отображение сообщает нам, что происходит с точками поверхности, когда они в первый раз на нее возвращаются, если, конечно, возвращаются. Это может показаться не ахти каким достижением, но такой подход снижает размерность пространства — число переменных в задаче. А это почти всегда хорошо.

Значение великолепной идеи Пуанкаре становится понятно, когда мы переходим к следующему по сложности типу решения — комбинации нескольких периодических движений. Вот простой пример такого движения: Земля обходит вокруг Солнца примерно за 365 дней, а Луна обходит вокруг Земли примерно за 27 дней. Так что движение Луны совмещает в себе эти два разных периода. Разумеется, весь смысл задачи трех тел заключается в том, что это описание не совсем точно, но «квазипериодические» решения такого рода часто встречаются в задачах с участием многих тел. Сечение Пуанкаре помогает распознать квазипериодические решения: когда они возвращаются к интересующей нас поверхности, то не попадают в точности в ту же точку, но точка, в которую они попадают раз за разом, крохотными шажочками обходит *на поверхности* замкнутую кривую.

Пуанкаре понял, что если бы все решения были такими, то можно было бы подобрать подходящий ряд и смоделировать их количественно. Но, проанализировав топологию отображения первого возвращения, он заметил, что все может быть куда сложнее. Две конкретные кривые, связанные динамикой, могут пересечься. Само по себе это не слишком плохо, но если вы пройдете по кривым до того места, где они вновь вернутся на нашу поверхность, то результирующие кривые вновь должны будут пересечься, но в другом месте. Проведите их еще круг, и они снова пересекутся. Мало того: эти новые кривые, полученные передвижением первоначальных кривых, на самом деле не новы. Они представляют собой части первоначальных кривых. Чтобы разобраться в этой топологии, потребовалось немало размышлений — ведь никто раньше подобными играми не занимался. В результате получается очень сложная картина, напоминающая сеть, сплетенную каким-то безумцем: кривые в ней ходят зигзагами туда-обратно, пересекая друг друга, а зигзаги эти сами, в свою очередь, ходят зигзагами туда-обратно и т. д. до любого уровня сложности. В конце концов, Пуанкаре заявил, что зашел в тупик:

> «Когда пытаешься описать фигуру, образованную этими двумя кривыми и их бесконечными пересечениями, каждое из которых соответствует дважды асимптотическому решению, то эти пересечения образуют своего рода сеть, паутину или бесконечно тонкое сито… Поражает сложность этой фигуры, которую я даже не пытаюсь нарисовать».

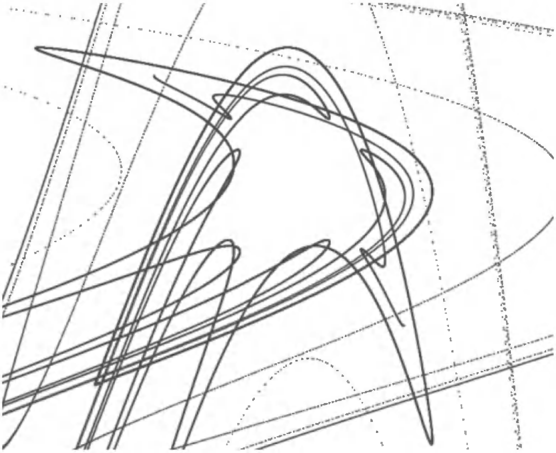


Рис. 31. Часть гомоклинного плетения. Полная картина была бы бесконечно сложной

Сегодня мы называем его картину гомоклинным («замкнутым на себя») плетением (см. рис. 31). Благодаря новым топологическим идеям, высказанным в 1960-е гг. Стивеном Смейлом, мы сегодня видим в этой структуре старого друга. Главное, что она помогла нам понять, — это то, что динамика *хаотична*. Хотя в уравнениях нет выраженного элемента случайности, их решения очень сложны и нерегулярны. В чем-то они похожи на по-настоящему случайные процессы. К примеру, существуют орбиты — более того, к этому типу относится большинство орбит, — движение которых в точности имитирует многократное случайное бросание монетки. Открытие того факта, что детерминистская система (т. е. система, будущее которой всецело и однозначно определяется ее текущим состоянием) может тем не менее обладать случайными чертами — замечательное достижение, оно изменило многие области науки. Мы уже не можем считать, что простые правила порождают простое поведение. Речь идет о том, что в обиходе часто называют теорией хаоса, и все это восходит непосредственно к Пуанкаре и его работе на приз короля Оскара.

Ну, почти все. На протяжении многих лет историки математики рассказывали об этом именно так. Но примерно в 1990 г. Джун Бэрроу-Грин обнаружила в недрах Института Миттага-Леффлера в Стокгольме печатный экземпляр работы Пуанкаре; пролистав его, она поняла, что он отличается от того варианта, который можно обнаружить в бесчисленных математических библиотеках по всему миру. Это оказалась официальная пояснительная записка к заявке Пуанкаре на приз, и в ней была ошибка. Подавая работу на конкурс, Пуанкаре упустил из виду хаотические решения. Он заметил ошибку прежде, чем работа была опубликована, доработал ее, выведя все, что было необходимо, — а именно хаос, — и заплатил (надо сказать, больше, чем стоил приз) за то, чтобы оригинальная версия была уничтожена, а в печать пошел исправленный вариант. Но по какой-то причине в архиве Института Миттага-Леффлера сохранился экземпляр первоначально ошибочной версии, хотя сама история забылась, пока Бэрроу-Грин не откопала ее и не опубликовала свое открытие в 1994 г.

Пуанкаре, судя по всему, считал, что хаотические решения несовместимы с разложениями в ряд, но это тоже оказалось ошибкой. Прийти к такому выводу было несложно: ряды казались слишком регулярными, чтобы представлять хаос, — на это способна только топология. Хаос — это сложное поведение, определяемое простыми правилами, так что это умозаключение небесспорно, но структура задачи трех тел определенно не допускает простых решений того рода, которые Ньютон вывел для двух тел. Задача двух тел интегрируема. Это означает, что в уравнениях достаточно сохраняющихся величин, таких как энергия, импульс и момент импульса, для однозначного определения орбиты. «Сохраняющихся» означает, что эти величины не меняют своего значения при движении тел по своим орбитам. Но задача трех тел неинтегрируема.

При всем том решения в виде рядов существуют, однако они не универсальны. Они не годятся для начальных состояний с нулевым моментом импульса — мерой суммарного вращения. Такие состояния бесконечно редки, поскольку нуль — всего лишь одно число среди бесконечного количества действительных чисел. Более того, в этих рядах фигурирует не время как таковое, а корень кубический из времени. Все это выяснил в 1912 г. финский математик Карл Фритьёф Зундман. Нечто аналогичное верно даже для задачи *n* тел опять же с редкими исключениями. Такой результат получил в 1991 г. Ван Цюдун. Но для системы из четырех или более тел у нас нет никаких достоверных данных о том, при каких именно обстоятельствах ряд не сходится, и мы никак не можем классифицировать эти обстоятельства. Мы знаем, однако, что такая классификация должна получиться очень сложной, потому что существуют решения, в которых все тела убегают в бесконечность или через некоторый конечный промежуток времени начинают колебаться с бесконечной частотой (см. главу 12). Физически такие решения — следствие нашего допущения, что все тела представляют собой точки, хотя и массивные. Математически они подсказывают нам, где искать самые дикие варианты поведения системы.

Серьезный успех в решении задачи *n* тел был достигнут для того частного случая, когда все тела обладают одинаковой массой. Такое допущение нечасто работает в небесной механике, но вполне разумно для некоторых неквантовых моделей элементарных частиц. А главный интерес такая постановка вопроса представляет, конечно же, для математиков. В 1993 г. Кристофер Мур нашел решение задачи трех тел для случая, когда все тела гоняются друг за другом по одной и той же орбите. Удивительна форма орбиты: это восьмерка, показанная на рис. 32. Несмотря на то что у орбиты есть точка самопересечения, тела никогда не сталкиваются.

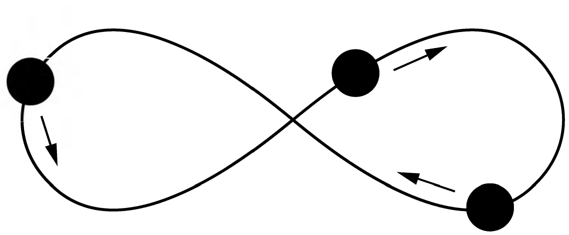


Рис. 32. Хореография на орбите-восьмерке

Расчет Мура был численным и проводился на компьютере. В 2001 г. Ален Ченсинер и Ричард Монтгомери заново независимо открыли это решение. Для этого они, с одной стороны, воспользовались давно известным в классической механике принципом наименьшего действия, а с другой — привлекли весьма хитроумную топологию, чтобы доказать, что такое решение существует. Орбиты тел периодичны во времени: через определенный временной промежуток все тела возвращаются к первоначальным позициям и скоростям, а затем повторяют те же движения до бесконечности. Для любой заданной суммарной массы существует по крайней мере одно такое решение для любого периода.

В 2000 г. Карлес Симо провел численный анализ и получил указания на стабильность восьмерки, за исключением, возможно, очень медленного долгосрочного дрейфа, известного как диффузия Арнольда и связанного с мелкими особенностями геометрии отображения карты возвращений Пуанкаре. При стабильности такого рода *почти* любые возмущения приводят объекты на орбиту, очень близкую к первоначальной, а среди мелких возмущений доля именно таких приближается к 100%. При тех редких возмущениях, при которых стабильность все же нарушается, орбита дрейфует от своего первоначального положении чрезвычайно медленно. Результат Симо вызвал удивление, поскольку в задаче трех тел равной массы стабильные орбиты встречаются редко. Численные расчеты показывают, что стабильность сохраняется даже в том случае, когда массы тел слегка различаются. Так что вполне возможно, что где-то во Вселенной три звезды с почти идентичными массами бесконечно преследуют одна другую на орбите в форме восьмерки. По оценке Дугласа Хегги, сделанной в 2000 г., число таких тройных звезд лежит между одной на галактику и одной на Вселенную.

Для орбиты в форме восьмерки характерна интересная симметрия. Возьмем для начала три тела A, B и C. Пройдем с ними треть орбитального периода и обнаружим тела на тех же позициях с теми же скоростями, как в начальный момент, только на тех же местах будут находиться соответственно тела B, C и A. После двух третей периода там же мы найдем тела C, A и B. Через полный период мы увидим в точности первоначальную картину. Решение такого рода известно как хореография — танец планет, в котором они через определенные промежутки времени меняются местами. Численные данные свидетельствуют о существовании хореографий в системах более чем трех тел: на рис. 33 представлены некоторые примеры таких систем. Сам Симо, в частности, отыскал огромное количество хореографий.

Но даже здесь многие вопросы остаются без ответа. У нас до сих пор нет строгого доказательства существования хореографий. Для систем более чем из трех тел все они представляются нестабильными. Скорее всего, так и есть, но это тоже надо доказать. Орбита в виде восьмерки для трех тел заданной массы при заданном периоде представляется единственной, но доказательства тому опять же нет, хотя в 2003 г. Томаш Капела и Петр Згличинский опубликовали компьютерное доказательство того, что она локально единственна — ни одна

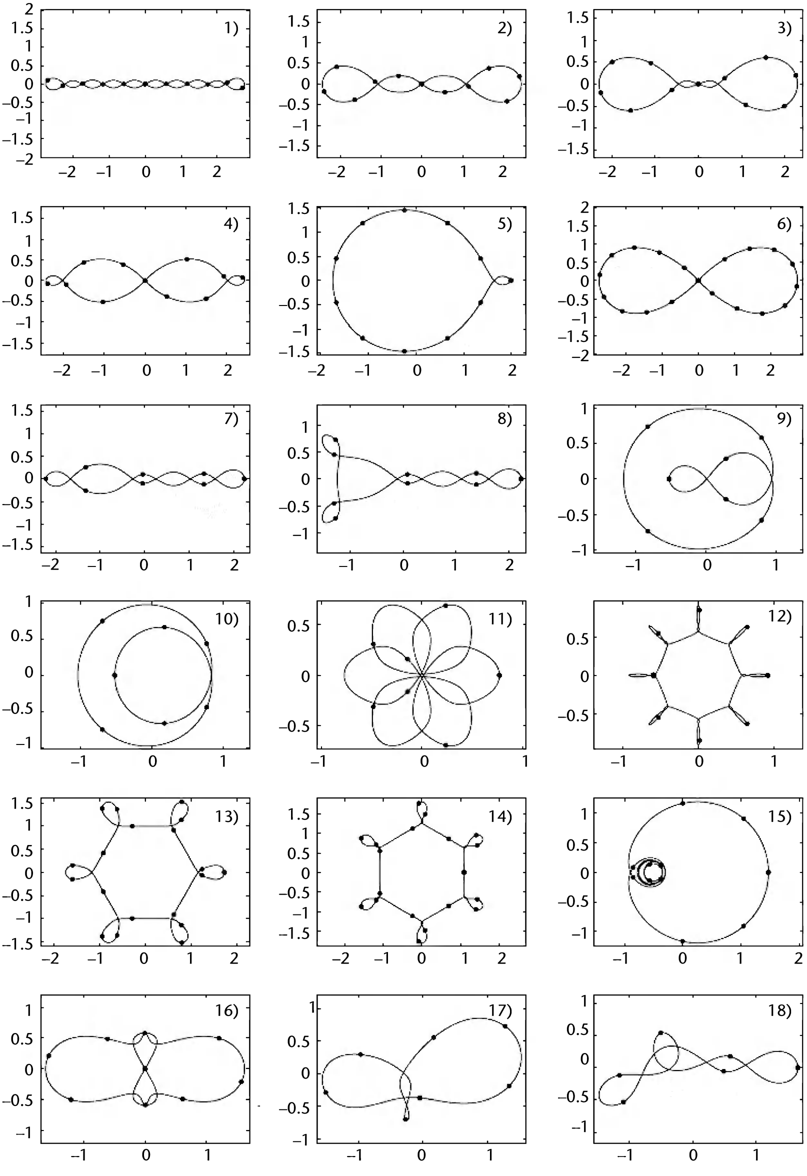


Рис. 33. Примеры хореографий

из близлежащих орбит не работает. Возможно, хореографии — это зерно еще одной великой задачи.

Итак, стабильна ли Солнечная система?

Может, да, а может, и нет.

Продолжая исследовать великое озарение Пуанкаре — возможность существования хаоса, — мы сегодня гораздо лучше разбираемся в теоретических вопросах, связанных с достижением стабильности. Оказалось, что это тонкая и сложная задача. К тому же она, как ни смешно, практически никак не связана с существованием или отсутствием решений в виде рядов. Работа Юргена Мозера и Владимира Арнольда позволила доказать, что различные упрощенные модели Солнечной системы стабильны почти при любых начальных состояниях, за исключением, возможно, эффекта диффузии Арнольда, который не допускает более сильных форм стабильности почти во всех задачах такого рода. В 1961 г. Арнольд доказал, что идеализированная модель Солнечной системы стабильна в этом смысле, но только при допущении, что планеты обладают чрезвычайно малыми массами по сравнению с массой центральной звезды, что их орбиты очень близки к круговым и находятся почти в одной плоскости. Если говорить о строгом доказательстве, то «почти» и «очень близки» здесь означает «различаются не более чем на 10^{-43} долю», и даже в этом случае точная формулировка гласит, что вероятность нестабильности равна нулю. Там, где речь идет о возмущениях, результаты часто бывают гораздо шире, чем то, что удается строго доказать, так что из всего этого следует, что планетная система, в разумной степени близкая к идеальной, вероятно, стабильна. Однако в нашей Солнечной системе допуски составляют 10^{-3} по массе и 10^{-2} по степени приближения к окружности и наклонению. Понятно, что это несколько больше, чем 10^{-43}, так что о применимости результатов Арнольда речь может идти лишь чисто теоретически. Тем не менее приятно, что в этом вопросе *хоть о чем-то* можно говорить определенно.

Практические стороны подобных задач тоже прояснились благодаря развитию мощных численных методов приближенного решения уравнений при помощи компьютера. Вообще-то это тонкий вопрос, ведь у хаоса есть одно важное свойство: маленькие ошибки способны очень быстро вырасти и погубить все решение целиком. Наши теоретические представления о хаосе и об уравнениях, подобных уравнениям Солнечной системы, в которой отсутствует трение, привели к развитию численных методов, свободных от многих наиболее неприятных свойств хаоса. Их называют симплектическими интеграторами. С их помощью удалось выяснить, что орбита Плутона хаотична. Однако это не означает, что Плутон беспорядочно носится по всей Солнечной системе, разрушая все вокруг себя. Это означает, что через 200 млн лет Плутон по-прежнему будет находиться где-то поблизости от своей нынешней орбиты, но где именно — мы не имеем ни малейшего представления.

В 1982 г. в рамках проекта Longstop под руководством Арчи Роя на суперкомпьютере проводилось моделирование внешних планет (начиная с Юпитера). В них не обнаружилось крупных нестабильностей, хотя некоторые планеты получали энергию за счет других планет странными путями. С тех пор две исследовательские группы занимаются развитием использовавшихся в проекте вычислительных методов и применением их к различным задачам, касающимся нашей Солнечной системы. Руководят этими группами Джек Уиздом и Жак Ласкар. В 1984 г. группа Уиздома предсказала, что спутник Сатурна Гиперион, вместо того чтобы вращаться честь по чести, должен беспорядочно кувыркаться, и последующие наблюдения подтвердили этот факт. В 1988 г. эта же группа в сотрудничестве с Джерри Сассменом построила собственный компьютер, спроектированный специально для работы с уравнениями небесной механики. В сущности, это цифровая модель Солнечной системы (в отличие от обычной механической модели, где движение планет — шариков на палочках — имитируется при помощи штырьков и шестеренок). Первый же расчет, смоделировавший следующие 845 млн лет Солнечной системы, вскрыл хаотичную природу Плутона. На данный момент группа Уиздома и ее последователи успели смоделировать динамику Солнечной системы на следующие несколько миллиардов лет.

Группа Ласкара опубликовала первые результаты по долгосрочному поведению Солнечной системы в 1989 г. При этом в расчетах использовалась усредненная форма уравнений, восходящих еще к Лагранжу. Понятно, что в таком расчете некоторые мелкие подробности размываются и исключаются из рассмотрения. Расчеты группы показали, что положение Земли на орбите хаотично, почти как у Плутона: если мы измерим сегодняшнее положение нашей планеты и ошибемся на 15 м, то ее положение на орбите через 100 млн лет невозможно будет предсказать сколько-нибудь определенно.

Единственный способ снизить влияние хаоса состоит в многократном моделировании с чуть разными начальными данными. Это позволяет получить спектр возможных вариантов вместе с вероятностью каждого из них. В 2009 г. Ласкар и Микаэль Гастино применили эту методику к Солнечной системе, рассмотрев 2500 различных сценариев. Различия между ними чрезвычайно малы — к примеру, это может быть сдвиг положения Меркурия на 1 м. Примерно в 1% вариантов будущего Меркурий становится нестабильным: он сталкивается с Венерой, рушится на Солнце или выбрасывается за пределы системы.

В 1999 г. Норман Мюррей и Мэтью Холман исследовали несоответствие результатов Арнольда и др., указывающих на стабильность, и моделирования, указывающего на нестабильность. «В чем дело? — спрашивали они. — Может быть, неверны численные результаты, а может быть, классические расчеты здесь неприменимы?» Воспользовавшись аналитическими, а не численными методами, они продемонстрировали, что классические расчеты применять нельзя. Возмущения, необходимые для отражения реальности, слишком велики.

Главный источник хаоса в Солнечной системе — близкое к резонансному состояние системы Юпитера, Сатурна и Урана и еще одной системы — Сатурна, Урана и Нептуна, хотя ее близость к резонансу не столь важна. Для проверки этого положения они использовали численные методы; получилось, что горизонт предсказания — мера времени, за которое небольшие ошибки приобретут достаточные масштабы, чтобы вызвать серьезные последствия, — составляет приблизительно 10 млн лет[28]. Их моделирование показывает, что Уран иногда опасно сближается с Сатурном, поскольку эксцентриситет его орбиты меняется хаотически и существует вероятность, что когда-нибудь он будет вообще выброшен прочь из Солнечной системы. Однако вероятное время такого события наступит через 10^{18} лет. Солнце взорвется и превратится в красный гигант гораздо раньше, примерно через 5 млрд лет. Это событие, естественно, скажется на всех планетах, не в последнюю очередь потому, что само Солнце при этом потеряет 30% массы. Земля отодвинется прочь от Солнца и, возможно, сумеет избежать захвата необычайно расширившимся светилом. Однако в настоящее время считается, что приливные взаимодействия со временем все же затянут Землю внутрь Солнца, а океаны нашей планеты вскипят и испарятся задолго до этого. Но поскольку типичная продолжительность жизни вида, с эволюционной точки зрения, не превышает 5 млн лет, нам вряд ли стоит беспокоиться обо всех этих потенциальных катастрофах. Мы погибнем гораздо раньше от каких-нибудь других причин.

При помощи этих же методов можно исследовать прошлое Солнечной системы: берем те же уравнения и пускаем время назад — простой математический фокус. До недавнего времени астрономы склонны были считать, что планеты всегда находились примерно на нынешних своих орбитах, — с тех самых пор, как сконденсировались из газопылевого облака, окружавшего зарождающееся Солнце. Более того, на основании их состава и формы орбит делались выводы о размерах и составе того самого первичного газопылевого облака. Сегодня же ученые склоняются к мнению, что планеты начинали свое существование вовсе не на нынешних орбитах. По мере того как из пылевого облака под действием внутреннего тяготения образовывались планеты, Юпитер — самая массивная из них — начал выстраивать остальные тела, да и сами они постоянно влияли друг на друга. Такую гипотезу предложили в 1984 г. Хулио Фернандес и Винг-Хуен Ип, но какое-то время их работу рассматривали скорее как любопытную, но незначительную диковинку. В 1993 г. Рену Малхотра всерьез задумался о том, как изменения в орбите Нептуна могли влиять на остальные планеты-гиганты. К нему присоединились другие исследователи, и постепенно проявилась картина чрезвычайно динамичной юности нашей Солнечной системы.

Планеты продолжали формироваться, и пришло время, когда Юпитер, Сатурн, Уран и Нептун были уже почти готовы, но между ними циркулировало громадное количество скальных и ледяных планетезималей — небольших тел около 10 км в поперечнике. После этого эволюция Солнечной системы шла путем миграции и столкновения планетезималей. Многие из них были выброшены в пространство, что снизило суммарную энергию и момент импульса четырех планет-гигантов. Поскольку все эти миры обладали разными массами и находились на разных расстояниях от Солнца, то и реагировали они по-разному. Нептун стал одним из победителей в орбитальной схватке за энергию и в результате отошел подальше от светила. Уран и Сатурн сделали то же самое, но в меньшей степени. Юпитер же в смысле энергии остался в проигравших и сместился внутрь системы. Но он был столь массивен, что далеко не ушел.

Остальные, меньшие тела Солнечной системы, тоже испытали на себе действие этих перемен. Текущее состояние системы, вроде бы стабильное, возникло в результате затейливого танца гигантов, в ходе которого разыгравшийся хаос бросил мельчайшие тела навстречу друг другу. Так стабильна ли Солнечная система? Вероятно, нет, но человечеству не удастся убедиться в этом на практике.

9

Закономерности простых чисел

Гипотеза Римана

В главе 2 мы рассматривали индивидуальные свойства простых чисел, и я сравнил их с зачастую непоследовательным и непредсказуемым поведением людей. Но люди обладают свободой воли, они могут принимать решения, исходя из своих соображений. А простые числа делают то, что подсказывает им логика арифметики, хотя нередко создается впечатление, что они тоже обладают собственной волей. Их поведение управляется странными совпадениями и часто лишено какой бы то ни было разумной структуры.

Тем не менее в мире простых чисел не правит анархия. В 1835 г. Адольф Кетле поразил современников, обнаружив математические закономерности в мире социальных явлений, которые зависят от сознательных решений разных людей или вмешательства судьбы: в мире рождений, свадеб, смертей, самоубийств. Закономерности были статистическими и касались не отдельных людей, а усредненного поведения больших человеческих масс. Именно так статистики извлекают порядок из индивидуальной свободы воли. Примерно в то же время математики начали осознавать, что такой фокус можно проделать и с простыми числами. Пусть каждое из них в отдельности — ярый индивидуалист, все вместе они подчиняются закону. Существуют скрытые закономерности.

Статистические закономерности проявляются тогда, когда мы рассматриваем сразу все множество простых чисел. К примеру: сколько простых чисел содержится в натуральном ряду до некоего определенного предела? На этот вопрос очень сложно ответить точно, но существуют прекрасные аппроксимации, и чем выше предел, тем точнее становятся приближенные значения. Иногда можно добиться, чтобы разница между приближенным и точным ответами была очень мала, но, как правило, это означало бы хотеть слишком многого. Большинство приближений в этой области являются асимптотическими. Это означает, что отношение приближенного значения к точному можно сделать очень близким к 1. При этом, хотя ошибка в процентах стремится к нулю, абсолютная ошибка может быть сколь угодно велика.

Если вы пытаетесь понять, как такое может быть, представим последовательность чисел — для некоего трудного для понимания свойства простых чисел — состоящую из степеней 100:

$$100,\ 10\,000,\ 1\,000\,000,\ 100\,000\,000.$$

Но реальные числа при этом таковы:

$$101,\ 10\,010,\ 1\,000\,100,\ 100\,001\,000,$$

т. е. лишняя единица сдвигается на каждом шаге влево на одну позицию. В этом случае отношение соответствующих чисел, чем дальше, тем ближе подходит к 1, а вот разность между ними приобретает вид:

$$1,\ 10,\ 100,\ 1000$$

и может достигать сколь угодно больших значений. Подобное поведение наблюдается в тех случаях, когда ошибка — разность между точным и приближенным ответом — беспредельно растет, но медленнее, чем растут сами числа.

Поиск асимптотических формул, имеющих отношение к простым числам, вдохновил математиков на создание новых методов теории чисел, основанных не на целых числах, а на комплексном анализе. Анализ — это строгое описание дифференциального и интегрального исчисления, включающего, как явствует из названия, два ключевых аспекта. В первом из них — дифференциальном исчислении — речь идет о скорости, с которой некая величина, называемая функцией, растет по отношению к другой величине. К примеру, положение тела зависит от времени, и скорость, с которой это положение изменяется со временем, представляет собой мгновенную скорость тела. Второй аспект — интегральное исчисление — имеет дело с расчетом площадей, объемов и тому подобных величин путем складывания большого числа очень маленьких кусочков. Процесс этот называется интегрированием. Примечательно, что интегрирование — это операция, обратная дифференцированию. Первоначальные формулировки Ньютона и Готфрида Лейбница требовали некоторых маневров с бесконечно малыми величинами, в связи с чем возникали вопросы о логической обоснованности этой теории. Со временем ученые разобрались с этими концептуальными вопросами, определив понятие предела — величины, к которой можно приблизиться на сколь угодно малое расстояние, но которой зачастую невозможно достичь. Именно в таком виде, в более строгих формулировках, метод получил название анализа.

Во времена Ньютона и Лейбница величины, о которых шла речь, представляли собой действительные числа, и результатом их работы, соответственно, стал действительный анализ. Когда же комплексные числа завоевали признание математиков, методы анализа естественным образом распространили и на них. Получился комплексный анализ, оказавшийся необычайно красивым и мощным инструментом. Вообще, когда дело доходит до анализа, комплексные функции ведут себя намного лучше, чем действительные. У них, конечно, есть свои особенности, но преимущества работы с комплексными функциями многократно перевешивают все их недостатки.

В какой-то момент математики с удивлением обнаружили, что арифметические свойства целых чисел можно с большой пользой переформулировать в терминах комплексных функций. До этого две системы ставили перед учеными очень разные вопросы и требовали использования очень разных методов. Но сегодня при помощи комплексного анализа — мощнейшего набора методик — можно открывать особые свойства функций теории чисел, а из них, в свою очередь, можно извлекать асимптотические формулы и многое другое.

В 1859 г. немецкий математик Бернхард Риман взял давнюю идею Эйлера и развил ее совершенно по-новому, определив так называемую дзета-функцию. Одним из результатов этой работы стала *точная* формула для количества простых чисел до заданного предела. Формула представляла собой бесконечную сумму, но специалистам по анализу к этому не привыкать. И это не было бесполезной игрой ума: благодаря этой формуле удалось получить новые подлинные знания о мире простых чисел. Мешала только одна маленькая неувязка. Хотя Риман мог доказать, что его формула точна, самые важные потенциальные следствия из нее полностью зависели от одного простого утверждения, касающегося дзета-функции, и вот это-то простое утверждение Риман никак не мог доказать. И сегодня, полтора столетия спустя, мы все еще не сумели сделать это. Сегодня это утверждение называется гипотезой Римана и представляет собой, по сути, священный Грааль чистой математики.

В главе 2 мы видели, что простые числа обыкновенно встречаются тем реже, чем они больше. Поскольку казалось, что точных формул для их распределения наверняка не существует, возникало естественное желание поискать статистические закономерности. В 1797–1798 гг. Лежандр подсчитал, сколько простых чисел помещается в натуральном ряду вплоть до различных пределов. Для этого он воспользовался таблицами простых чисел, которые незадолго до того составили Георг Вега и Антон Фелькель. Веге, судя по всему, вообще нравились сложные расчеты: он составил таблицы логарифмов и в 1789 г. стал обладателем мирового рекорда по вычислению числа π, которое он посчитал до 140-го десятичного знака (из них 126 были посчитаны верно). А Фелькелю просто нравилось искать простые числа. Его главная работа вышла в 1776 г. и называлась «Таблица всех простых делителей чисел до 10 000 000, за исключением тех, что делятся на 2, 3 или 5». Для проверки делимости на 2, 3 и 5 есть простые способы, упомянутые в главе 2, и он сэкономил в книге много места, опустив эти числа. Лежандр открыл эмпирическую приближенную формулу для количества простых чисел, меньших заданного числа x, и обозначил это количество π (x). Если вы привыкли воспринимать π только как символ для обозначения числа 3,14159, это потребует привыкания, но в любом контексте несложно понять, что именно имелось в виду, даже если вы не заметили, что символы даны в несколько разном начертании. В 1808 г. в тексте Лежандра по теории чисел утверждалось, что значение $\pi(x)$, судя по всему, очень близко к значению выражения $x/(\log x - 1{,}08366)$.

В 1849 г. в письме к астроному Иоганну Энке Гаусс сообщил, что в свое время, лет в 15, сделал на полях таблицы логарифмов запись, в которой утверждалось количество простых чисел, меньших или равных x, составляет $x/\log x$ для больших x. Гаусс не опубликовал это наблюдение (как и многие другие свои открытия), возможно, потому, что не имел доказательства. В 1838 г. Дирихле сообщил Гауссу об аналогичной приближенной формуле, найденной им самим. По существу, эта формула сводится к логарифмической интегральной функции[29]

$$\mathrm{Li}(x) = \int_0^x \frac{\mathrm{d}t}{\log t}.$$

По мере того как x становится большим, отношение $\mathrm{Li}(x)$ к $x/\log x$ стремится к 1; это означает, что если одно из них асимптотически равно $\pi(x)$, то асимптотически равно и второе, но рис. 34 позволяет предположить (совершенно верно), что $\mathrm{Li}(x)$ — лучшее приближение, чем $x/\log x$. Точность $\mathrm{Li}(x)$ впечатляет. К примеру,

$$\pi(1\,000\,000\,000) = 50\,847\,534,$$
$$\mathrm{Li}(1\,000\,000\,000) = 50\,849\,234{,}9.$$

Аппроксимация в виде $x/\log x$ хуже: в данном случае ее значение 48 254 942,4.

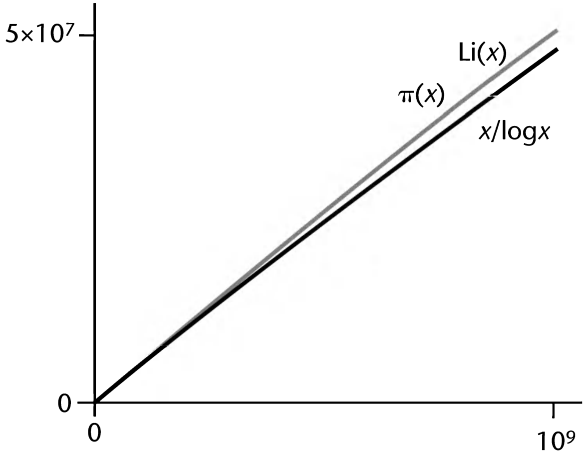


Рис. 34. В данном масштабе графики $\pi(x)$ и $\mathrm{Li}(x)$ (серые) неразличимы. Однако значения $x/\log x$ (черный график) заметно меньше. Здесь x откладывается на горизонтальной оси, а значение функции — на вертикальной оси

Приближенная формула с использованием $\mathrm{Li}(x)$ или $x/\log x$ стала известна как теорема о распределении простых чисел, где слово «теорема» использовалось в смысле «предположение». Поиск доказательства того, что эти формулы асимптотичны к $\pi(x)$, стал одной из ключевых открытых задач теории чисел. Многие математики пытались одолеть ее при помощи традиционных методов этой области науки, и некоторые подошли к ответу достаточно близко, однако всегда оставалась какая-то хитрая посылка, которую никак не удавалось доказать. Нужны были новые методы. Они появились в результате любопытного переформулирования двух древних, еще евклидовых, теорем о простых числах.

Теорема о распределении простых чисел была ответом на евклидову теорему о том, что простые числа уходят в бесконечность и могут быть сколь угодно большими. Другая фундаментальная евклидова теорема говорит о единственности разложения на простые множители: каждое положительное целое число есть произведение простых чисел, причем *только одного их набора*. В 1737 г. Эйлер понял, что первую теорему можно переформулировать в виде поразительной формулы из действительного анализа, и тогда второе утверждение становится простым следствием этой формулы. Для начала я представлю формулу, а затем попытаюсь разобраться в ней. Вот она:

$$\frac{1}{1-2^{-s}}\times\frac{1}{1-3^{-s}}\times\ldots\times\frac{1}{1-p^{-s}}\times\ldots=$$
$$=\frac{1}{1^s}+\frac{1}{2^s}+\frac{1}{3^s}+\frac{1}{4^s}+\frac{1}{5^s}+\frac{1}{6^s}+\frac{1}{7^s}+\ldots$$

Здесь p принимает все простые значения, а s — константа. Эйлера интересовал в основном случай, при котором s — целое число, но его формула работает и для действительных чисел, в случае если s больше единицы. Это условие необходимо для того, чтобы ряд в правой части сошелся, т. е., будучи продолжен до бесконечности, принял бы осмысленное значение.

Это необыкновенная формула. В левой части мы перемножаем бесконечно много выражений, которые зависят только от простых чисел. В правой — складываем бесконечное число выражений, которые зависят от всех положительных целых чисел. Эта формула выражает, на языке анализа, некоторое отношение между целыми и простыми числами. Главное отношение такого рода — это единственность разложения на простые множители, именно она оправдывает существование формулы.

Я кратко опишу основной этап, чтобы показать, что за всем этим стоит разумная идея. Воспользовавшись школьной алгеброй, мы можем разложить выражение в ряд по p. Этот ряд напоминает правую часть формулы, но включает только степени p. А именно:

$$\frac{1}{1-p^{-s}}=\frac{1}{1^s}+\frac{1}{p^s}+\frac{1}{p^{2s}}+\frac{1}{p^{3s}}+\dots$$

Когда мы перемножим все эти ряды, для всех простых p, и раскроем все скобки, мы получим комбинации с любыми степенями простых чисел, т. е. с любыми целыми положительными степенями. Все они выглядят как величины, обратные (т. е. единица, деленная на) s-й степени данного числа, и все возникают лишь единожды в связи с единственностью разложения на простые множители. Таким образом, получаем ряд в правой части.

Никому до сих пор не удалось найти простой алгебраической формулы для суммы этого ряда, хотя формул с интегралами немало. Поэтому мы присвоили ей особый символ — греческую букву дзета (ζ) — и определили новую функцию:

$$\zeta(s)=\frac{1}{1^s}+\frac{1}{2^s}+\frac{1}{3^s}+\frac{1}{4^s}+\frac{1}{5^s}+\frac{1}{6^s}+\frac{1}{7^s}+\dots$$

Вообще говоря, Эйлер не использовал символ ζ и рассматривал только положительные целые значения s, но я буду и дальше называть приведенный выше ряд эйлеровой дзета-функцией. Воспользовавшись своей формулой, Эйлер заключил, что существует бесконечно много простых чисел (для этого он рассматривал значения s, близкие к единице). Главной его целью было получить формулы вроде $\zeta\ (2) = \pi^2/6$ и найти сумму ряда для четных целых s. Развивать свою революционную идею дальше он не стал.

Другие математики заметили упущение Эйлера и рассмотрели нецелые значения s. В двух работах 1848 и 1850 гг. русский математик Пафнутий Чебышев предложил великолепную идею: попытаться доказать теорему о распределении простых чисел при помощи анализа. Начал он со связи между простыми числами и математическим анализом, обеспечиваемой эйлеровой дзета-функцией. Он не добился полного успеха, поскольку считал s действительным числом, а аналитические возможности действительного анализа весьма ограничены. Зато он сумел доказать, что для больших x отношение $\pi(x)$ к $x/\log x$ лежит между двумя константами, одна из которых чуть больше единицы, а вторая — чуть меньше. Это был уже реальный результат, хотя пока еще не такой, как хотелось. Он позволил Чебышеву доказать постулат Бертрана, предложенный в 1845 г.: если взять любое натуральное число (≥ 2) и удвоить его, то между двумя этими числами обязательно найдется простое число.

Вот теперь сцена была готова к появлению Римана. Он тоже понял, что дзета-функция — это ключ к теореме о распределении простых чисел, но для реализации этого подхода ему пришлось предложить смелое расширение: определить дзета-функцию не только действительной, но и комплексной переменной. А начать можно с ряда Эйлера. Он сходится для любых действительных s больше единицы, и если использовать для комплексного s в точности ту же формулу, то ряд будет сходиться при любых s, у которых действительная часть больше 1. Однако Риман обнаружил, что можно сделать и лучше. Применив процедуру так называемого аналитического продолжения, он расширил определение $\zeta(s)$ на *все* комплексные числа, за исключением 1. Это значение s исключено потому, что при $s = 1$ значение дзета-функции становится бесконечным[30].

В 1859 г. Риман собрал все свои мысли о дзета-функции в одну статью, заголовок которой можно перевести как «О количестве простых чисел, не превышающих заданной величины». В ней он привел полную и точную формулу $\pi(x)$[31]. Я опишу более простую формулу, эквивалентную римановой, чтобы показать, как появляются нули дзета-функции. Идея заключается в том, чтобы подсчитать, сколько простых чисел, или степеней простых чисел, укладывается до любого заданного предела. Однако вместо того чтобы сосчитать каждое число по одному разу, как функция $\pi(x)$ делает с простыми числами, мы придаем большим простым числам дополнительный вес. Более того, любая степень простого числа учитывается в соответствии с логарифмом этого простого числа. Так, для предела 12 мы имеем следующие степени простых чисел:

$$2, 3, 4 = 2^2, 5, 7, 8 = 2^3, 9 = 3^2, 11,$$

поэтому взвешенный подсчет дает

$$\log 2 + \log 3 + \log 2 + \log 5 + \log 7 + \log 2 + \log 3 + \log 11,$$

что составляет примерно 10,23.

Воспользовавшись методами анализа, информацию об этом более хитроумном способе подсчета простых чисел можно превратить в информацию об обычном способе. Однако этот метод приводит к более простым формулам, и присутствие логарифма — не слишком дорогая цена за это. В этих терминах точная формула Римана говорит о том, что взвешенный подсчет до предела x эквивалентен

$$-\sum_{\rho}\frac{x^{\rho}}{\rho}+x-\frac{1}{2}\log\left(1-x^{-2}\right)-\log 2\pi,$$

где Σ обозначает сумму по всем числам ρ, для которых $\zeta(\rho)$ равна нулю, исключая отрицательные четные целые числа. Эти значения называются нетривиальными нулями дзета-функции. Тривиальные нули — это отрицательные четные целые числа –2, –4, –6… Во всех этих точках дзета-функция равняется нулю из-за формулы, которая используется в определении аналитического продолжения, но, как выяснилось, для римановой формулы эти нули несущественны (как и почти везде в других местах).

На случай, если формула вас немного пугает, я укажу главное: хитрый способ подсчета простых чисел до заданного предела x, который при помощи кое-каких аналитических фокусов можно превратить в обычный способ, *в точности* эквивалентен сумме по всем нетривиальным нулям дзета-функции простого выражения x^ρ/ρ *плюс некая несложная функция от* x. Если вы специалист по комплексному анализу, вы сразу увидите, что доказательство теоремы о распределении простых чисел эквивалентно доказательству того, что взвешенный подсчет до предела x асимптотически сходится к x. Воспользовавшись комплексным анализом, получим: это утверждение верно, если у всех нетривиальных нулей дзета-функции действительная часть лежит между 0 и 1. Чебышев не смог этого доказать, но подошел достаточно близко, чтобы извлечь полезную информацию.

Почему нули дзета-функции так важны? Одна из базовых теорем комплексного анализа утверждает, что при некоторых формальных условиях функция комплексной переменной полностью определяется значениями переменной, при которых функция равна нулю или бесконечности, плюс некоторая дополнительная информация о поведении функции в этих точках. Эти особые точки известны как нули и полюсы функции. В действительном анализе эта теорема не работает — и это одна из причин, по которым комплексный анализ завоевал такую популярность, несмотря на необходимость извлекать корень квадратный из –1. У дзета-функции один полюс (при $s = 1$), так что все ее характеристики определяются нулями (если, конечно, не забывать о существовании этого единственного полюса).

Для удобства Риман работал в основном с зависимой кси-функцией $\xi(x)$, которая тесно связана с дзета-функцией и получается из метода аналитического продолжения. Он заметил:

> «Весьма вероятно, что все [нули кси-функции] действительны. Хотелось бы, конечно, иметь строгое доказательство этого факта, но после нескольких бесплодных попыток я отложил поиск такого доказательства, поскольку этого не требуется для непосредственных целей моего исследования».

Это заявление о кси-функции эквивалентно аналогичному заявлению о зависимой от нее дзета-функции. А именно: все нетривиальные нули дзета-функции представляют собой комплексные числа вида: они лежат на *критической линии* «действительная часть равна 1/2» (см. рис. 35). Эта версия замечания и есть знаменитая гипотеза Римана.

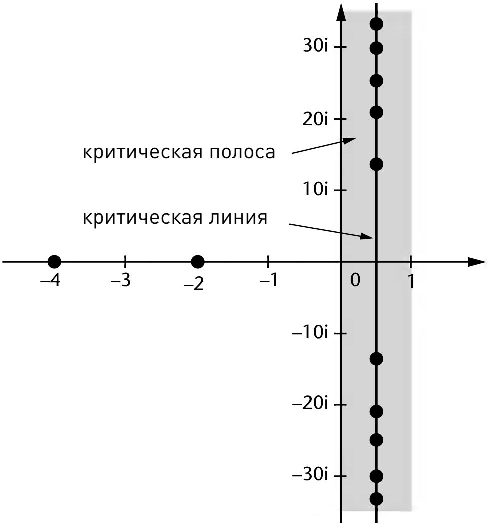


Рис. 35. Нули дзета-функции, критическая линия и критическая полоса

Замечание Римана звучит достаточно небрежно, как будто высказано между делом и эта гипотеза не имеет особого значения. И это действительно так, если говорить только о программе Римана по доказательству теоремы о распределении простых чисел. Но во многих других вопросах верно обратное. Многие считают гипотезу Римана важнейшим из остающихся на сегодняшний день открытыми математических вопросов.

Чтобы понять, почему это так, мы должны последовать за рассуждениями Римана чуть дальше. В тот момент ученый был нацелен на теорему о распределении простых чисел. Его точная формула предлагала верный путь к этому достижению: нужно было разобраться в нулях дзета-функции или эквивалентной ей кси-функции. Полная риманова гипотеза для этого не нужна, достаточно доказать, что у всех нетривиальных нулей дзета-функции действительная часть лежит в промежутке от 0 до 1, т. е. что сами комплексные корни лежат на расстоянии не более 1/2 от римановой критической линии — в так называемой критической полосе. Это свойство нулей подразумевает, что сумма по всем нулям дзета-функции, фигурирующая в приведенной выше точной формуле, представляет собой конечную константу. Асимптотически для больших x она вообще может потеряться. Единственный член формулы, который сохранит свое значение при очень больших x, это сам x. Все остальные сложные слагаемые асимптотически пропадают в сравнении с x. Следовательно, взвешенная сумма асимптотически стремится к x, и это доказывает теорему о распределении простых чисел. Так что, по иронии судьбы, роль нулей дзета-функции заключается в том, чтобы доказать, что они не вносят существенного вклада в точную формулу.

Риман так и не довел свою программу до логического конца. Более того, он никогда больше ничего не писал по этому вопросу. Но два других математика, приняв у него эстафету, показали, что догадка Римана верна. В 1896 г. Жак Адамар и Шарль-Жан де ла Валле Пуссен независимо друг от друга вывели теорему о распределении простых чисел, доказав, что все нетривиальные нули дзета-функции лежат в пределах критической полосы. Доказательства у обоих получились очень сложными и техничными, но тем не менее свою задачу они выполнили. Возникла новая мощная область математики — аналитическая теория чисел. Применение ей нашлось в самых разных уголках теории чисел: с ее помощью решали давние задачи и выявляли новые закономерности. Другие математики позже нашли несколько более простых доказательств теоремы о числе простых, а Атле Сельберг и Пал Эрдеш открыли даже очень сложное доказательство, вовсе не требовавшее применения комплексного анализа. Но к тому моменту при помощи идеи Римана было доказано бесчисленное множество важных теорем, включая аппроксимации многих функций теории чисел. Так что это новое доказательство хоть и добавило в эту историю каплю иронии, но ни на что, в сущности, не повлияло. В 1980 г. Дональд Ньюман нашел гораздо более простое доказательство, для которого достаточно оказалось всего лишь одной из самых базовых теорем комплексного анализа — теоремы Коши.

Хотя Риман объявил свою гипотезу ненужной для достижения ближайших целей, оказалось, что она жизненно необходима для разрешения многих других вопросов теории чисел. Прежде чем обсуждать гипотезу Римана, нам стоит взглянуть на некоторые теоремы, которые — если бы гипотеза была доказана — из нее следуют.

Одно из важнейших следствий — это величина погрешности в теореме о распределении простых чисел. Теорема, как вы помните, утверждает, что для большого x отношение $\pi(x)$ к $\mathrm{Li}(x)$ приближается к 1, причем чем дальше, тем сильнее. Иными словами, разница между двумя функциями снижается до нуля *относительно величины* x[32]. Однако реальная разница при этом может расти (и растет). Просто она делает это медленнее, чем растет сам x. Компьютерные расчеты позволяют предположить, что величина погрешности примерно пропорциональна $\quad x \log x$. Если гипотеза Римана верна, это утверждение можно доказать. В 1901 г. Хельге фон Кох доказал, что гипотеза Римана логически эквивалентна оценке

$$\pi(x) - \mathrm{Li}(x) \leq \frac{1}{8\pi} \quad x \log x$$

для всех $x \geq 2657$. Здесь вертикальными линиями обозначена абсолютная величина: разность, умноженная на ± 1, чтобы сделать ее положительной. Эта формула дает наилучшие возможные ограничения для разницы между $\pi(x)$ и $\mathrm{Li}(x)$.

Из гипотезы Римана можно получить немало других оценок для функций теории чисел. К примеру, из нее прямо следует, что сумма делителей n меньше

$$e^{\gamma} n \log \log n$$

для всех $n \geq 5040$, где γ — постоянная Эйлера ($\gamma = 0{,}57721$)[33]. Эти утверждения могут показаться случайными и странными фактами, но хорошая оценка для важной функции жизненно важна во многих приложениях, и большинство специалистов по теории чисел отдали бы свою правую руку ради того, чтобы доказать любую из них.

Кроме того, гипотеза Римана говорит нам, насколько велико может быть расстояние между последовательными простыми числами. Типичный размер промежутка между ними можно вывести на основании теоремы о распределении простых чисел: в среднем промежуток между простым числом p и следующим простым числом сравним с $\log p$. Некоторые промежутки могут быть меньше, некоторые больше, но математикам жилось бы легче, если бы можно было сказать наверняка, насколько велики могут быть самые большие из них. Харальд Крамер доказал в 1936 г., что если гипотеза Римана верна, то промежуток при простом числе p не может превышать величины $p \log p$, домноженной на некую константу.

Но подлинное значение гипотезы Римана куда глубже. Существуют далеко идущие обобщения и сильное подозрение, что тот, кто сумеет доказать гипотезу Римана, сможет, вероятно, доказать и связанную с ней обобщенную гипотезу Римана. А это, в свою очередь, даст математикам власть над обширными областями теории чисел.

Обобщенная гипотеза Римана вырастает из более подробного описания простых чисел. Все простые числа, кроме двойки, нечетные, и в главе 2 мы видели, что все нечетные простые можно разделить на два типа: те, что на 1 больше числа, кратного 4, и те, что на 3 больше числа, кратного 4. Говорят, что это числа вида $4k + 1$ или $4k + 3$, где k — число, на которое вы умножаете 4, чтобы получить данное простое число. Приведем короткий список первых нескольких простых чисел того и другого типа, вместе с соответствующими числами, кратными 4:

кратное 4	0	4	8	12	16	20	24	28	32	36
плюс 1	—	5	—	13	17	—	—	29	—	37
плюс 3	3	7	11	—	19	23	—	—	—	—

Прочерки указывают на то, что соответствующее число не простое.

Сколько существует простых чисел того и другого типа? Как они распределены среди всех простых чисел или среди всех целых чисел? Евклидово доказательство того факта, что простых чисел существует бесконечно много, можно без больших усилий модифицировать, доказав при этом, что существует бесконечно много простых чисел вида $4k + 3$. Доказать, что простых чисел вида $4k + 1$ тоже бесконечно много, гораздо сложнее, — это можно сделать, но лишь при помощи некоторых достаточно сложных теорем. Разница в подходах обусловлена тем, что любое число вида $4k + 3$ имеет делитель того же вида, а в отношении чисел вида $4k + 1$ это не всегда верно.

В числах этих двух видов нет ничего чудесного или священного. Все простые числа, кроме 2 и 3, имеют вид $6k + 1$ или $6k + 5$, и мы можем задать в отношении них аналогичные вопросы. Если уж на то пошло, все простые числа, кроме 5, имеют вид $5k + 1$, $5k + 2$, $5k + 3$, $5k + 4$. Мы оставляем в стороне числа вида $5k$, поскольку они кратны 5 и, соответственно, все, кроме 5, не являются простыми.

Кстати говоря, на любой из подобных вопросов нетрудно выдвинуть разумное предположение — простые числа в арифметической последовательности. Случай с $5k$ достаточно типичен. Эксперимент быстро показывает, что числа приведенных выше четырех видов имеют примерно равные шансы оказаться простыми. Вот похожая таблица:

кратные 5	5	10	15	20	25	30	35	40
плюс 1	—	11	—	—	—	31	—	41
плюс 2	7	—	17	—	—	—	37	—
плюс 3	—	13	—	23	—	—	—	43
плюс 4	—	—	19	—	29	—	—	—

Так что должно существовать бесконечное количество простых чисел каждого вида, и в среднем к каждому виду должна относиться четверть всех простых чисел до заданного предела.

Для некоторых видов доказать, что простых чисел такого вида существует бесконечно много, совсем несложно. Для других видов требуются более изощренные рассуждения. Но до середины XIX в. никому не удавалось доказать, что существует бесконечно много простых чисел каждого возможного вида, не говоря уже о том, чтобы доказать их более или менее равномерное распределение. Лагранж в 1785 г. в работе, посвященной закону квадратичной взаимности — глубокому свойству квадратов простых модулей, — принимал этот факт без доказательства. Результаты дали очевидно полезные следствия, и пора было кому-нибудь это доказать. В 1837 г. Дирихле выяснил, как применить идеи Эйлера, связанные с теоремой о распределении простых чисел, для доказательства обоих этих утверждений. Первым делом следовало определить аналоги дзета-функции для этих типов простых чисел. То, что получилось, называется L-функциями Дирихле. К примеру, в случае $4k + 1/4k + 3$ возникает следующая функция:

$$\mathrm{L}\,(s, \chi) = 1 - 3^{-s} + 5^{-s} - 7^{-s} + 9^{-s} - \ldots,$$

где коэффициенты равны +1 для чисел вида $4k + 1$, –1 для чисел вида $4k + 3$ и 0 для остальных. Греческую букву χ называют характером Дирихле, и это напоминает нам о том, какие именно знаки следует использовать.

Для римановой дзета-функции важен не только ряд, но и его аналитическое продолжение, придающее функции значения во всех комплексных точках. То же относится и к L-функции, и Дирихле определил подходящее аналитическое продолжение. Приспособив к случаю идеи, которые использовались для доказательства теоремы о распределении простых чисел, он сумел доказать аналогичную теорему о простых числах особых видов. К примеру, число простых чисел вида $5k + 1$, меньших или равных x, асимптотически приближается к $\mathrm{Li}(x)/4$; то же относится и к остальным трем случаям $5k + 2$, $5k + 3$, $5k + 4$. Это означает, что простых чисел каждого вида бесконечно много.

Риманова дзета-функция — это особый случай L-функции Дирихле для простых чисел вида $1k + 0$, т. е. для всех простых чисел. Обобщенная гипотеза Римана представляет собой очевидное обобщение оригинальной гипотезы: нули любой L-функции Дирихле либо имеют действительную часть, равную $1/2$, либо являются тривиальными нулями, действительная часть которых отрицательна или больше единицы.

Если обобщенная гипотеза Римана верна, то верна и обычная его гипотеза. Многие следствия обобщенной гипотезы Римана аналогичны следствиям обычной. К примеру, схожие границы ошибки можно доказать для аналогичных версий теоремы о распределении простых чисел в применении к простым числам любого конкретного вида. Однако обобщенная гипотеза Римана подразумевает много такого, что совершенно отличается от всего, что мы можем вывести из обычной гипотезы Римана. Так, в 1917 г. Годфри Харди и Джон Литтлвуд доказали, что из обобщенной гипотезы Римана следует гипотеза Чебышева, в том смысле, что (буквально) простые числа вида $4k + 3$ встречаются чаще, чем числа вида $4k + 1$. Согласно теореме Дирихле, оба вида равновероятны в конечном итоге, но это не мешает простым числам вида $4k + 3$ выигрывать у чисел $4k + 1$, конечно, в правильной игре.

У обобщенной гипотезы Римана есть также важные следствия, имеющие отношение к проверке на простоту, такие как тест Миллера 1976 г., упомянутый в главе 2. Если обобщенная гипотеза Римана верна, то тест Миллера дает нам эффективный алгоритм проверки. Оценка эффективности более поздних тестов тоже зависит от обобщенной гипотезы Римана. Существуют и важные приложения для алгебраической теории чисел. Помните, в главе 7 говорилось, что новое определение идеальных чисел Куммера, данное Дедекиндом, привело к рождению новой фундаментальной концепции — понятия идеала. Разложение на простые множители в кольцах алгебраических целых чисел существует, но может не быть единственным. Разложение идеалов на простые множители работает много лучше: и существование, и единственность гарантированы. Так что имеет смысл заново рассмотреть все вопросы о множителях в терминах идеалов. В частности, существует понятие «простого идеала» — разумной и удобной аналогии простого числа.

Зная это, естественно спросить, есть ли у эйлеровой связи между обычными простыми числами и дзета-функцией аналог для простых идеалов. Если да, то весь мощный аппарат аналитической теории чисел применим к алгебраическим числам. Оказывается, это можно сделать, с глубокими и очень серьезными последствиями. Результат — дзета-функция Дедекинда — по одной такой функции на каждую систему алгебраических чисел. Существует глубокая связь между комплексными аналитическими свойствами дедекиндовой дзета-функции и арифметикой простых чисел в соответствующей системе алгебраических целых чисел. И, разумеется, существует аналог гипотезы Римана: все нетривиальные нули дедекиндовой дзета-функции лежат на критической линии. Понятие «обобщенная гипотеза Римана» теперь включает в себя и это утверждение.

Даже генерализация — еще не конец истории дзета-функции. Она вдохновила ученых на определение аналогичных функций в нескольких других областях математики — от абстрактной алгебры до теории динамических систем. Во всех этих областях существуют еще более масштабные аналоги гипотезы Римана. Некоторые из них даже доказаны. В 1974 г. Пьер Делинь доказал такой аналог для многообразий над конечными полями. Обобщения, известные как дзета-функции Сельберга, тоже удовлетворяют аналогу гипотезы Римана. То же можно сказать о дзета-функции Госса. Однако существуют другие обобщения — дзета-функции Эпштейна, для которых аналог гипотезы Римана неверен. Здесь бесконечное множество нетривиальных нулей лежит на критической линии, но некоторые — нет, что продемонстрировал Эдвард Титчмарш. С другой стороны, эти дзета-функции не имеют эйлеровой формулы в виде произведения и потому не похожи на римановы дзета-функции в аспекте, который вполне может оказаться принципиально важным.

Имеется множество косвенных свидетельств того, что гипотеза Римана — как оригинальная, так и обобщенная — справедлива. Много хорошего следовало бы из истинности этих гипотез. Ни одно из этих следствий за все время не удалось опровергнуть, а ведь сделать это — то же самое, что опровергнуть гипотезу Римана. Но ни доказательства, ни опровержения пока нет. Широко распространено мнение, что доказательство оригинальной гипотезы Римана открыло бы дорогу и к доказательству обобщенного ее варианта. Но на самом деле, возможно, лучше было бы атаковать сразу обобщенную гипотезу Римана во всей ее грозной красе — воспользоваться всем арсеналом доступных на сегодняшний день методов, доказать, а затем вывести оригинальную гипотезу Римана как ее частный случай.

В пользу гипотезы Римана имеется также огромное количество экспериментальных данных — по крайней мере огромное на первый взгляд, пока кто-нибудь не плеснет холодной воды, чтобы остудить горячие головы. По данным Карла Людвига Зигеля, Риман вычислил несколько первых нулей своей дзета-функции, но не стал публиковать результат. Они находятся в точках

$$\frac{1}{2} \pm 14{,}135i,\ \frac{1}{2} \pm 21{,}022i,\ \frac{1}{2} \pm 25{,}011i.$$

Нетривиальные нули всегда располагаются парами, как здесь. Я написал в них, а не 0,5, потому что действительная часть в этих случаях известна точно, выяснена при помощи общих результатов комплексного анализа и известных свойств дзета-функции. То же можно сказать и о компьютерных расчетах, о которых речь пойдет ниже. Они не просто показывают, что нули находятся очень близко к критической линии; они действительно находятся на ней.

В 1903 г. Йорген Грам продемонстрировал численно, что первые десять нулей (т. е. ±-пар) лежат на критической линии. К 1935 г. Титчмарш увеличил число таких нулей до 195. В 1936 г. Титчмарш и Лесли Комри доказали, что первая 1041 пара нулей лежит на критической линии. Это был последний раз, когда подобные расчеты проводились вручную.

Алан Тьюринг больше всего известен тем, что во время войны работал в Блетчли-парке, где участвовал в разгадывании германского кода «Энигма», а также своими работами, заложившими фундамент компьютерных вычислений и искусственного интеллекта. Но, помимо всего этого, Тьюринг интересовался и аналитической теорией чисел. В 1953 г. он открыл более эффективный способ вычисления нулей дзета-функции и определил при помощи компьютера, что первые 1104 пары нулей лежат на критической линии. Свидетельства того, что все нули до некоторого предела лежат на критической линии, множились и множились. Нынешний рекорд, полученный Янником Саутером и Патриком Демишелем в 2004 г., составляет 10 трлн (10^{13}). Тем временем математики и компьютерщики проверяли другие диапазоны нулей. На сегодня все без исключения нетривиальные нули, когда-либо кем-либо рассчитанные, лежат на критической линии.

Все это может показаться исчерпывающим доказательством, но математики не спешат принимать его на веру, и не без причины. Может показаться, что 10 трлн — это очень много, но в теории чисел часто значение имеет не само число, а его логарифм, а он пропорционален числу знаков в числе. Натуральный логарифм от 10 трлн чуть меньше 30. Мало того, во многих задачах фигурирует логарифм от логарифма или даже логарифм от логарифма от логарифма. В этих терминах 10 трлн — это *крохотная* величина, так что численное доказательство до 10 трлн включительно почти ничего не значит.

Существуют и кое-какие обобщенные аналитические доказательства, к которым эта критика не относится. Харди и Литтлвуд доказали, что на критической линии лежит бесконечное число нулей. Другие математики показали в точном смысле, что почти все нули лежат очень близко к критической линии. Сельберг доказал, что ненулевая доля нулей лежит непосредственно на критической линии. Норманн Левинсон доказал, что эта доля — по крайней мере треть и теперь она увеличена по крайней мере до 40%. Все эти результаты позволяют предположить, что если гипотеза Римана неверна, то нули, не лежащие на критической линии, очень велики и встречаются очень редко. К несчастью, главное следствие из всего этого заключается в том, что если такие исключения существуют, то найти их будет необычайно трудно.

Но зачем волноваться? Ведь численных свидетельств должно быть достаточно, чтобы убедить любого разумного человека? К несчастью, нет. Численные свидетельства не убеждают математиков, и в данном случае это не просто педантизм и придирки: они действуют разумно. В математике в целом, а особенно в теории чисел, обширные, на первый взгляд, «экспериментальные» данные часто имеют гораздо меньший вес, чем может показаться.

Наглядным примером служит гипотеза Пойа, которую в 1919 г. выдвинул венгерский математик Дьердь Пойа. Он предположил, что по крайней мере половина всех целых чисел вплоть до заданной величины имеет нечетное число простых множителей. Повторяющиеся множители в данном случае учитываются отдельно, а начинаем мы с 2. К примеру, число простых множителей для предела 20 приведено в табл. 2, где последняя колонка отражает процент чисел до данного предела с нечетным числом простых множителей.

Все значения в последней колонке выше 50%, а более обширные расчеты позволяют предположить, что это всегда так. В 1919 г., без всяких компьютеров, исследователи не смогли найти чисел, которые опровергли бы эту гипотезу. Но в 1958 г. Брайан Хазелгроув доказал при помощи аналитической теории чисел, что гипотеза Пойа неверна для некоего числа — числа, не превосходящего 1,845 × 10^{361}, если быть точным. Как только на сцене появились компьютеры, Шерман Леман показал, что гипотеза неверна для 906 180 359. К 1980 г. Минору Танака доказал, что минимальное из таких чисел 906 150 257. Так что вы могли бы собрать экспериментальные данные по всем числам почти до миллиарда и не понять, что гипотеза неверна.

Тем не менее приятно знать, что число 906 150 257 необычайно интересно.

Разумеется, сегодняшние компьютеры, если их как следует запрограммировать, опровергли бы гипотезу Пойа в несколько секунд. Но иногда не помогают даже они. Классический пример — число Скьюза, где первоначально громадное количество численных данных указывало на то, что некая знаменитая гипотеза верна, но на самом деле она неверна. Это гигантское число появилось в задаче, тесно связанной с гипотезой Римана: аппроксимацией $\pi(x)$ функции $\mathrm{Li}(x)$. Как мы только что видели, теорема о распределении простых чисел утверждает, что, когда x становится большим, отношение этих двух величин стремится к 1. Численные расчеты указывают на более сильное утверждение: это отношение всегда меньше 1, т. е. $\pi(x)$ меньше $\mathrm{Li}(x)$. В 2008 г. численные расчеты Тадея Котника показали, что это верно для x меньше 10^{14}. К 2012 г. Дуглас Столл

Таблица 2. Процент чисел до заданного предела, имеющих нечетное число простых множителей

Число	**Разложение на простые множители**	**Число простых среди множителей**	**Доля простых, %**
2	2	1	100
3	3	1	100
4	2^2	2	66
5	5	1	75
6	2×3	2	60
7	7	1	66
8	2^3	3	71
9	3^2	2	62
10	2×5	2	55
11	11	1	60
12	$2^2 \times 3$	3	63
13	13	1	66
14	2×7	2	61
15	3×5	2	57
16	2^4	4	53
17	17	1	56
18	2×3^2	3	58
19	19	1	61
20	$2^2 \times 5$	3	63

и Демишель повысили этот предел до 10^{18}, и такой же результат независимо от них получил Андрей Кульша. А расчеты Томаша Оливейра-и-Сильва позволяют предположить, что предел может быть увеличен до 10^{20}.

Звучит, кажется, исчерпывающе. Данные здесь сильнее, чем лучшие численные результаты, полученные до сих пор для гипотезы Римана. Но в 1914 г. Литтлвуд доказал, что эта гипотеза неверна — и как доказал! По мере того как x проходит через положительные действительные значения, разность $\pi(x) - \mathrm{Li}(x)$ меняет знак (с отрицательного на положительный или наоборот) *бесконечно часто*. В частности, $\pi(x)$ *больше* $\mathrm{Li}(x)$ для некоторых достаточно больших значений x. Однако доказательство Литтлвуда ничего не говорило о конкретных значениях x.

В 1933 г. его ученик, южноафриканский математик Стенли Скьюз, оценил, насколько большим должен быть x: не более 10∧10∧10∧34, где знак ∧ обозначает «возвести в степень». Это настолько гигантское число, что если все его цифры напечатать в книге — довольно скучной книге, состоящей из 1 с бесконечными нулями, — то Вселенная не вместила бы этой книги, даже если бы каждая цифра была размером с элементарную частицу. Более того, чтобы доказательство работало, Скьюзу пришлось принять на веру истинность гипотезы Римана. К 1955 г. он нашел способ обойтись без гипотезы Римана, но не бесплатно: его оценка увеличилась до 10∧10∧10∧963.

Эти числа слишком велики даже для прилагательного «астрономический», но дальнейшие исследования помогли снизить их до величин, которые уже можно охарактеризовать как космологические. В 1966 г. Леман заменил числа Скьюза на 10^{1165}. Те Риеле в 1987 г. понизил эту оценку до 7×10^{370}, а в 2000 г. Картер Бейз и Ричард Хадсон свели ее к $1{,}39822 \times 10^{316}$. Затем Чжоу Куок Фай и Роджер Плаймен срезали еще немножко и довели ограничение до $1{,}39801 \times 10^{316}$. Это изменение может показаться несущественным, но на самом деле данная оценка меньше предыдущей на 2×10^{312}. А Саутер и Демишель еще улучшили этот результат, сведя его к $1{,}3971667 \times 10^{316}$.

Но пока суд да дело, в 1941 г. Аурел Уинтнер доказал, что маленькая, но ненулевая доля целых чисел удовлетворяет неравенству $\pi(x) > \mathrm{Li}(x)$. В 2011 г. Столл и Демишель просчитали первые 200 млрд нулей дзета-функции, что позволяет судить о $\pi(x)$ для всех x вплоть до $10^{10\,000\,000\,000\,000}$, и нашли доказательство того, что если x меньше, чем $3{,}17 \times 10^{114}$, то $\pi(x)$ меньше $\mathrm{Li}(x)$. Так что для данной конкретной проблемы все свидетельства по крайней мере до 10^{18}, а очень может быть, что и до 10^{114} или даже больше, только вводят в заблуждение. Переменчивые боги теории чисел любят пошутить за счет людей.

Было предпринято немало попыток доказать или опровергнуть гипотезу Римана. На сайте Мэттью Уоткинса «Предлагавшиеся доказательства гипотезы Римана» перечислены около 50 таких попыток, сделанных уже после 2000 г. Во многих доказательствах найдены ошибки, и ни одно из них профессиональное сообщество не признало верным.

Одной из самых разрекламированных в последние годы стала попытка Луи де Бранжа, предпринятая в 2002 г. Он распространил среди математиков рукопись, в которой попытался доказать гипотезу Римана при помощи области анализа, имеющей дело с преобразованиями на пространствах бесконечной размерности и известной как функциональный анализ. У специалистов были основания принять попытку де Бранжа всерьез. Незадолго до того он так же распространил доказательство гипотезы Бибербаха о разложении в ряд комплексных функций. В первоначальном доказательстве обнаружились ошибки, но со временем было установлено, что основная его идея работает. Однако в данном случае казалось, что предложенный де Бранжем метод доказательства гипотезы Римана не имеет шансов на успех. Брайан Конри и Ли Сяньцзинь указали на некоторые непреодолимые, как пока представляется, препятствия.

Возможно, самая серьезная надежда на доказательство гипотезы Римана заключается в новых методах или радикально новых подходах к задаче. Как мы неоднократно видели, прорывы в работе по великим математическим задачам происходят, как правило, тогда, когда кому-нибудь удается связать давно известную задачу с совершенно другой, далекой от нее областью математики. Прекрасный пример — Великая теорема Ферма: как только ее удалось интерпретировать как вопрос об эллиптических кривых, прогресс не заставил себя ждать.

Сегодня тактика де Бранжа вызывает вопросы, но сам его подход стратегически вполне оправдан. Своими корнями он уходит в устное предположение, сделанное около 1912 г. Давидом

Гильбертом и независимо от него Дьердем Пойа. Эдмунд Ландау тогда спросил у Пойа, по каким таким физическим причинам должна быть верна гипотеза Римана. В 1982 г. Пойа вспоминал, что нашел-таки тогда ответ: нули дзета-функции следует связать с собственными значениями так называемого самосопряженного оператора. Речь идет о характеристических числах, связанных с особым типом преобразования. В квантовой физике — одной из важнейших областей применения — эти числа определяют энергетические уровни системы, и существует стандартная несложная теорема о том, что собственные числа этого особого типа преобразования всегда действительны. Если бы собственные числа некоего самосопряженного оператора совпадали с нулями кси-функции, то гипотеза Римана была бы несложным следствием этого факта. Пойа не стал публиковать эту идею — он не мог привести пример такого оператора, а пока примера нет — все это журавль в небе. Но в 1950 г. Сельберг доказал свою «формулу следа», которая связывает геометрию поверхности с собственными числами соответствующего оператора. Это сделало идею чуть более правдоподобной.

В 1972 г. Хью Монтгомери побывал в Институте перспективных исследований в Принстоне. В разговоре с физиком Фрименом Дайсоном он упомянул замеченные им некоторое время назад удивительные статистические свойства нетривиальных нулей дзета-функции. Дайсон сразу же отметил их сходство со статистическими свойствами случайных эрмитовых матриц — еще одного частного случая оператора, который используется для описания квантовых систем, таких как атомное ядро. В 1999 г. Ален Конн предложил формулу следа, аналогичную формуле Сельберга. Подтвердить ее — означало бы доказать обобщенную гипотезу Римана. В том же 1999 г. физики Майкл Берри и Йон Китинг предположили, что требуемый оператор может быть получен при квантовании одного хорошо известного понятия из классической физики, имеющего отношение к импульсу. Получившуюся в результате гипотезу Берри можно рассматривать как частную версию гипотезы Гильберта–Пойа.

Эти идеи, помогающие соотнести гипотезу Римана с глубинными областями математической физики, очень интересны. Они показывают, что решение может прийти из, казалось бы, никак не связанных с ней областей математики, и внушают надежду на то, что когда-нибудь вопрос с гипотезой Римана будет закрыт. Однако до сих пор они не привели к окончательному прорыву, и у нас нет оснований считать, что решение уже близко. Гипотеза Римана остается одной из самых запутанных и волнующих загадок во всей математике.

Сегодня у исследователей появился новый стимул к борьбе за доказательство гипотезы Римана: крупный приз.

В математике не существует Нобелевской премии. Самой престижной наградой в этой области является Филдсовская премия за выдающиеся открытия, вместе с которой вручается медаль. Эта премия названа в честь канадского математика Джона Филдса, который и завещал на нее средства. Раз в четыре года на Международном конгрессе математиков двум, трем или четырем молодым ученым не старше 40 лет вручают золотую медаль и денежную премию (в настоящее время это $15 000).

Многие представители математической науки считают правильным, что в их области не присуждается Нобелевская премия. В настоящее время она составляет чуть больше миллиона долларов, а такая сумма легко может исказить цели исследователей и породить споры о приоритетах. Однако отсутствие крупной математической премии также может исказить представления общества о значимости и полезности этой науки. Можно подумать, что открытия, за которые никто не хочет платить, не так уж важны. Возможно, поэтому не так давно появились две очень престижные новые математические премии. Одна из них — Абелевская — присуждается ежегодно Норвежской академией науки и словесности и названа в честь великого норвежского математика Нильса Хенрика Абеля. Вторая награда — это премии за решение семи «проблем тысячелетия», объявленные Математическим институтом Клэя. Этот институт основали в 1998 г. в Кембридже (штат Массачусетс) американский бизнесмен Лэндон Клэй и его жена Лавиния. Лэндон Клэй активно занимается паевыми инвестиционными фондами и при этом любит и уважает математику. Его организация проводит встречи, выделяет гранты на исследования, организует публичные лекции и присуждает ежегодную премию за математические исследования.

В 2000 г. сэр Майкл Атья и Джон Тейт, ведущие математики Великобритании и США, объявили, что Математический институт Клэя учредил новую премию, которая должна будет стимулировать работу над семью важнейшими нерешенными задачами математики. Эти задачи будут известны как «проблемы тысячелетия», а надлежащим образом опубликованное и отреферированное решение любой из них будет вознаграждено денежной суммой в $1 млн. Все вместе эти задачи призваны привлечь внимание к некоторым центральным для математики вопросам, до сих пор не имеющим ответов. Вопросы эти были тщательно отобраны лучшими математиками мира. Немалый приз должен ясно показать обществу: математика имеет огромную ценность. Всякий, кто имеет отношение к науке, прекрасно знает, что интеллектуальная ценность вполне может быть выше любых денег, но все же деньги помогают сосредоточиться. Самой известной и давней из задач тысячелетия является гипотеза Римана. Это единственный вопрос, который вошел одновременно и в список Гильберта (1900), и в список задач тысячелетия. Остальные шесть проблем тысячелетия обсуждаются далее в главах 10–15. Тем не менее математики не особенно гонятся за призами, и работа над гипотезой Римана продолжалась бы и без обещанной премии. Все, что для этого нужно, — новая перспективная идея.

Стоит также помнить о том, что гипотезы, даже освященные временем, иногда оказываются ошибочными. Сегодня большинство математиков, судя по всему, считает, что когда-нибудь гипотеза Римана будет доказана. Некоторые, однако, думают, что она, возможно, все-таки неверна, и где-то в дебрях очень больших чисел может скрываться нуль дзета-функции, который не лежит на критической линии. Если такой «контрпример» существует, то он, скорее всего, окажется очень-очень большим.

Однако на переднем крае математики просто мнение стоит немного. Интуиция зачастую очень помогает ученым, но известно немало случаев, когда это замечательное чувство ошибалось. Житейский здравый смысл может лгать, оставаясь при этом и общепризнанным, и здравым. Литтлвуд, один из лучших знатоков комплексного анализа, выразился вполне однозначно: в 1962 г. он сказал, что уверен в ошибочности гипотезы Римана, и добавил, что нет никаких мыслимых причин, по которым она была бы верна. Кто прав? Поживем, увидим.

10

Какой формы сфера?

Гипотеза Пуанкаре

Анри Пуанкаре был одним из величайших математиков конца XIX в., слегка эксцентричным, но исключительно прозорливым ученым. Он был членом французского Бюро долгот — организации, созданной для решения астрономических задач для целей навигации, слежения за временем и измерения Земли и других планет. Членство в ней привело Пуанкаре к мысли об установлении международной системы временны́х зон. Кроме того, оно вдохновило ученого на размышления о времени и, в частности, на предвосхищение некоторых открытий Эйнштейна в области теории относительности. Вклад Пуанкаре виден в математике всюду, от теории чисел до математической физики. В частности, он был одним из основателей топологии, математики непрерывных преобразований.

В 1904 г. Пуанкаре наткнулся на простой, казалось бы, вопрос и вдруг понял, что ответ на него, который он прежде использовал в работе как нечто очевидное, доказать не в состоянии. «Этот вопрос увел бы нас далеко в сторону», — написал он, погрешив против истины: на самом деле, этот вопрос упрямо отказывался вести его *куда бы то ни было*. Хотя сам Пуанкаре сформулировал задачу в виде вопроса, известность она приобрела как гипотеза Пуанкаре — все были уверены в том, что ответом на вопрос должно быть «да». Это еще одна из семи «проблем тысячелетия» по версии Института Клэя, что вполне справедливо, поскольку несложная на вид задача оказалась одной из самых сложных во всей топологии. Ответ на вопрос Пуанкаре дал в 2002 г. молодой русский математик Григорий Перельман. Его решение привнесло в науку массу новых идей и методов — так много, что математическому сообществу потребовалось несколько лет, чтобы «переварить» представленное доказательство и признать его верным.

За свое достижение Перельман был удостоен самой престижной в математике Филдсовской премии, но отказался от нее. Он избегает всякой публичности. Ему предложили миллион долларов — премию Института Клэя, — но и ее Перельман не принял. Деньги ему тоже не нужны. Он хотел лишь, чтобы его труд был принят математическим сообществом. Со временем так и произошло, но, к несчастью, процесс занял немало времени. Да и вообще, наивно было ожидать признания без публичности и премий. Но склонная к затворничеству натура Перельмана не смогла принять эти неизбежные следствия успеха.

Мы уже встречались с топологией в связи с теоремой о четырех красках, и я прибегал тогда к расхожему сравнению: «геометрия на резиновом листе». Евклидова геометрия имеет дело с прямыми линиями, окружностями, длинами и углами. Она разворачивается на плоскости или в пространстве трех измерений, где становится более сложной. Плоскость похожа на бесконечный лист бумаги, и у нее с бумагой есть одна общая черта: она не растягивается, не сжимается и не сгибается. Бумагу можно скатать в трубочку, и она может слегка съежиться или растянуться — особенно если пролить на нее кофе. Но невозможно обернуть бумагой шар так, чтобы на листе не образовалось складок. Математически евклидова плоскость — штука жесткая. В геометрии Евклида две фигуры — два треугольника, квадрата или круга — равны, если один из них получен из другого посредством жесткого перемещения. А «жесткость» перемещения означает, что расстояния при этом не меняются.

Но что если использовать вместо бумаги эластичный резиновый лист? Он, в отличие от бумаги, растягивается и сгибается, при некотором желании его можно даже сжать. Длины и углы на эластичном листе не имеют фиксированных значений. Более того, если лист *достаточно* эластичен, треугольников, квадратов и кругов на нем тоже нет. Можно деформировать треугольник на резиновом листе так, что у него появится дополнительный угол. Можно даже превратить его в круг (см. рис. 36). Какими бы понятиями ни оперировала геометрия на резиновом листе, ясно, что традиционным евклидовым концепциям в ней места нет.

Может показаться, что «геометрия на резиновом листе» будет настолько гибкой, что в ней вообще не найдется места для сколько-нибудь постоянных смыслов, а значит, и доказать ничего невозможно. Но это не так. Нарисуйте, к примеру, треугольник и поставьте внутри него точку. Если вы начнете растягивать и деформировать лист, превращая треугольник в круг, то одно свойство вашего чертежа все же сохранится: точка останется внутри. Согласен, теперь она находится внутри круга, а не треугольника, но это не важно: она же все равно не *снаружи*. Чтобы переместить точку наружу, вам придется разорвать лист, а это будет означать нарушение правил игры.

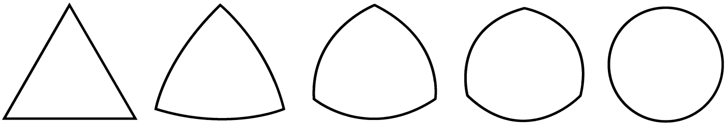


Рис. 36. Топологическое преобразование треугольника в круг

И еще одно свойство не изменяется при искажении. Треугольник — это простая замкнутая кривая. Это линия, замкнутая сама на себя, без свободных концов и самопересечений. Восьмерка — тоже замкнутая кривая, но уже не простая —

у нее есть самопересечение. При деформировании резинового листа треугольник может изменить форму, но обязательно останется простой замкнутой кривой. Невозможно, к примеру, превратить его в восьмерку, не разорвав листа.

В трехмерной топологии все пространство становится эластичным. Не так, как куб из резины, который, если снять давление, возвращается к первоначальной форме, а как гель, форму которого можно менять без всякого сопротивления. Топологическое пространство бесконечно эластично: можно взять участок размером с рисовое зернышко и раздуть его до размеров Солнца. Можно тянуть из него щупальца, пока он не начнет напоминать формой осьминога. Единственное, чего делать не разрешается, — это каким бы то ни было образом нарушать непрерывность. Не следует допускать разрывов пространства и вообще проделывать любые операции, способные разделить соседние точки.

Какие свойства пространственных фигур сохраняются при всех непрерывных деформациях? Не длина, не площадь, не объем… А вот заузленность сохраняется. Если завязать кривую узлом и соединить концы, создав замкнутую петлю, то узел уже никуда не денется. Как бы вы ни деформировали пространство, кривая останется завязанной. Таким образом, мы имеем дело с геометрией нового типа, где важные и осмысленные концепции кажутся, на первый взгляд, несколько расплывчатыми: «внутри», «замкнутый», «простой», «завязанный». Называется эта новая геометрия весьма респектабельно: топологией. Нематематику она может показаться странной или даже абсурдной, но на поверку это одна из основных областей математики XX в., и свое значение она сохраняет и в XXI в. А Пуанкаре — один из тех, кого мы в первую очередь должны за это благодарить.

История топологии началась почти за столетие до Пуанкаре — в 1813 г. Симон Люилье, швейцарский математик, при жизни не снискал громкой славы, но отверг крупную сумму денег, которую кто-то из его родственников предлагал ему за принятие церковного сана. Люилье предпочел карьеру в математике. Работал он в основном в тихой математической заводи: занимался теоремой Эйлера о многогранниках. В главе 4 мы упоминали один занятный и вроде бы ни с чем не связанный факт: если у многогранника F граней, V вершин и E ребер, то $F - E + V = 2$. Люилье большую часть жизни исследовал варианты этой формулы, и с высоты сегодняшнего дня ясно, что он сделал важнейший шаг в направлении топологии, когда обнаружил, что формула Эйлера не всегда верна. Ее применимость зависит от качественных характеристик многогранника.

Формула верна для многогранников без отверстий, которые можно нарисовать на поверхности сферы или на поверхности, полученной из сферы непрерывным преобразованием. Но если в многограннике есть отверстия, формула перестает работать. К примеру, рамка для картины, изготовленная из прямоугольного в сечении деревянного бруска имеет 16 граней, 32 ребра и 16 вершин; здесь $F - E + V = 0$. Люилье доработал формулу Эйлера для подобных экзотических многогранников: если в многограннике g отверстий, то $F - E + V = 2 - 2g$. Так был открыт первый важный топологический инвариант: величина, которая связана с пространством и не меняется при любых непрерывных преобразованиях пространства. Инвариант Люилье позволяет точно подсчитать, сколько отверстий имеет та или иная поверхность, не определяя строго, что такое «отверстие». Это полезно, поскольку «отверстие» — понятие достаточно сложное. Отверстие — не часть поверхности и не область вне поверхности. Очевидно, это свойство того, как поверхность располагается в окружающем пространстве. Но открытие Люилье показывает, что то, что мы интерпретируем как *количество* отверстий, есть свойство, изначально присущее поверхности и не зависящее от окружающего пространства. Нет необходимости определять отверстия, а затем считать их; лучше вообще не делать этого.

После Люилье следующей ключевой фигурой в предыстории топологии стал Гаусс. Работая в различных областях математики, он столкнулся с несколькими другими топологическими инвариантами. Работа в комплексном анализе, особенно над доказательством того, что каждое полиномиальное уравнение имеет по крайней мере одно решение в комплексных числах, заставила его рассмотреть порядок кривой на плоскости: сколько оборотов она делает относительно заданной точки. Задачи из области электричества и магнетизма подсказали коэффициент зацепления двух замкнутых кривых: сколько раз одна из них проходит сквозь другую. Эти и другие примеры привели Гаусса к мысли о существовании некоего пока не открытого раздела математики, в котором предлагался бы последовательный взгляд на качественные свойства геометрических фигур. Он ничего не публиковал по этой теме, но упоминал ее в письмах и рукописях.

Кроме того, он сообщил эти соображения своему ученику Иоганну Листингу и своему сотруднику Августу Мёбиусу. Я уже упоминал ленту Мёбиуса — поверхность, у которой есть лишь одна сторона и один край. Статью о ней он опубликовал в 1865 г., а саму ее можно увидеть на рис. 9 в главе 4. Мёбиус указал, что определение «имеющая одну сторону» интуитивно понятно, но тем не менее не точно, и предложил вместо него близкое свойство, которое можно определить совершенно строго. Это свойство — ориентируемость. Поверхность ориентируема, если ее можно покрыть сетью треугольников со стрелками вдоль сторон таким образом, что всюду, где два треугольника имеют общую сторону, стрелки указывают в противоположных направлениях. Если вы разместите такую сеть на плоскости и направите все стрелки в треугольниках, к примеру, по часовой стрелке, то именно так и произойдет. А вот на ленте Мёбиуса такая сеть невозможна.

Первая публикация Листинга по топологии появилась раньше, в 1847 г. Называлась она «Лекции по топологии», и это был первый текст, в котором использовалось это слово. Неформально Листинг пользовался им уже около 10 лет. Кроме того, тогда для той же цели использовалась латинская фраза *analysis situs,* т. е. «анализ размещения», но со временем она вышла из употребления. Книга Листинга не содержит ничего особенно значительного, но дает одно фундаментальное понятие: покрытие поверхности сетью треугольников. В 1861 г., за четыре года до Мёбиуса, Листинг описал ленту Мёбиуса и исследовал связность — вопрос о том, можно ли разбить пространство на две или более несвязанных областей. На основе работы Листинга другие математики, в том числе Вальтер фон Дик, провели полную топологическую классификацию поверхностей, считая их замкнутыми (не имеющими краев) и компактными (конечной протяженности). Оказалось, что любая ориентируемая поверхность топологически эквивалентна сфере с конечным числом g ручек (см. рис. 11 в середине и справа, глава 4). Число g называют родом поверхности, и именно его определяет инвариант Люилье. Если $g = 0$, это сфера, а если $g > 0$, мы имеем тор с g отверстиями. Аналогичный ряд поверхностей, начиная с простейшей неориентируемой поверхности — проективной плоскости, — образуют и неориентируемые поверхности. Этот метод был расширен и на поверхности с краями. Каждый край — это замкнутая петля, и единственное, что нужно знать дополнительно, это количество таких петель.

Гипотезу Пуанкаре легче понять, если рассмотреть для начала один из базовых методов, используемых при классификации поверхностей. Ранее я сравнил топологию с деформированием объекта, изготовленного из резины или геля, и подчеркнул, что преобразование обязательно должно быть *непрерывным*. По иронии судьбы, один из центральных методов топологии включает операцию, которая, на первый взгляд, нарушает непрерывность: разрезание объекта на кусочки. Однако непрерывность восстанавливается при помощи серии правил, описывающих, какие куски соединены друг с другом и как именно. Примером может служить то, как мы определили тор, отождествив противоположные стороны квадрата (см. рис. 12, глава 4).

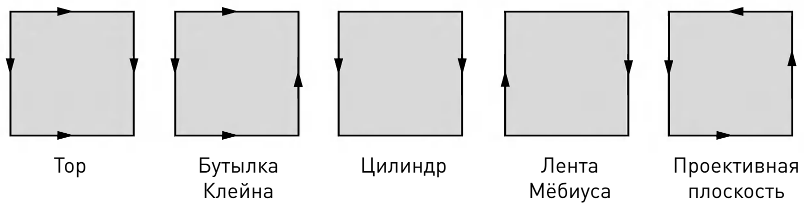


Рис. 37. Пять различных топологических пространств можно получить, по-разному отождествив противоположные края квадрата

Отождествление четко различимых точек позволяет нам представлять сложные топологические пространства при помощи простых составляющих. Квадрат — это всего лишь квадрат, и ничего больше, но с правилами отождествления квадрат может быть тором, бутылкой Клейна, цилиндром, лентой Мёбиуса или проективной плоскостью, в зависимости от характера правил (см. рис. 37). Так что когда я, объясняя непрерывное преобразование, сравнил его с растягиванием и сгибанием резинового листа, я, строго говоря, требовал больше необходимого. Нам разрешено также разрезать лист на промежуточной стадии при условии, что, в конце концов, мы либо соединим куски в точности так же, как было вначале, либо обозначим правила, которые позволят это сделать. С точки зрения тополога, *сформулировать* правило склеивания краев — это то же самое, что склеить их. Если, конечно, не забывать это правило в ходе дальнейших операций.

Классический метод классификации поверхностей начинается с рисования на поверхности сети треугольников. Затем мы делаем достаточно много разрезов вдоль сторон треугольника так, чтобы фигура развернулась в плоский многоугольник. Правила склеивания, определяемые тем, как мы делаем разрезы, подскажут нам, как отождествить разные края многоугольника, восстанавливая первоначальную поверхность. В этот момент вся интересующая нас топология заключена в правилах склеивания. Эта классификация доказывается алгебраической обработкой правил и превращением их в правила, определяющие тор с g отверстиями или одну из аналогичных ему неориентируемых поверхностей. У современной топологии есть и другие способы добиться того же самого, но и она часто пользуется техникой «разрезания и склеивания». Этот метод легко обобщается на пространства любой размерности, но он слишком ограничен, чтобы дать возможность классифицировать многоразмерные топологические пространства без дополнительной помощи.

Около 1900 г. Пуанкаре занимался тем, что развивал более раннюю свою работу по топологии поверхностей и разрабатывал значительно более общую методику, применимую к пространствам с любым числом измерений. Основной идеей его исследования был поиск топологических инвариантов: чисел или алгебраических формул, связанных с пространствами, которые при непрерывной деформации остаются неизменными. Если топологические инварианты двух пространств различны, то одно из них невозможно преобразовать в другое и, значит, они топологически различны.

Начал он с обобщения топологического инварианта Люилье $F - E + V$ на многомерные пространства, сделанного итальянским математиком Энрико Бетти в 1870 г. Сейчас оно, отчасти несправедливо, известно как эйлерова характеристика. Бетти заметил, что наибольшее число замкнутых кривых, которые можно нарисовать на поверхности рода g, не поделив ее при этом на несвязные куски, равняется $g - 1$. Это еще один способ топологически охарактеризовать поверхность. Бетти обобщил эту идею на «числа связности» любой размерности, которые Пуанкаре назвал числами Бетти, и этот термин используется до сих пор. Так, k-мерное число Бетти означает число k-мерных отверстий в пространстве.

Пуанкаре определил на основе чисел Бетти более чувствительный инвариант, получивший название гомологии. Для него характерна гораздо более четкая алгебраическая структура. Подробнее мы поговорим о гомологии в главе 15. Пока же достаточно сказать, что гомология анализирует наборы многомерных «граней» в подобной сети и задается вопросом о том, какие из них образуют границу топологического диска. Диск не имеет отверстий, в отличие от тора, так что мы можем быть уверены, что в пределах любого набора граней, образующего границу, отверстий нет. Напротив, мы можем обнаруживать отверстия путем разделения наборов граней на те, что образуют границу, и на те, что границы не образуют. Таким образом мы можем построить серию инвариантов пространства, известных как его гомологические группы. Слово «группа» здесь используется как термин из абстрактной алгебры, означающий, что из любых двух объектов группы при помощи операции, для которой выполняются несколько соответствующих алгебраических правил, может быть получен объект той же группы. Позже, когда нам потребуется это понятие, я расскажу о нем немного больше. Для каждого измерения от 0 до *n* существует одна такая группа, и для каждого пространства мы получаем серию топологических инвариантов со всевозможными интереснейшими алгебраическими свойствами.

Листинг классифицировал все топологические поверхности — пространства размерности 2. Очевидным следующим шагом было посмотреть на пространства размерности 3. И простейшим пространством для начала стала сфера, т. е. бесконечно тонкая поверхность шара. Внутренняя часть шара не считается частью сферы: это всего лишь особенность, возникающая вследствие вложения сферической поверхности в пространство. По существу, у нас есть только поверхность, топологически эквивалентная поверхности шара. Можно представить ее себе как пустотелый мяч с бесконечно тонкой оболочкой.

«Правильный» трехмерный аналог сферы, называемой трехмерной, — это *не шар*. Шар, конечно, трехмерен, но у него есть граница — его поверхность, сфера. Сама сфера границы не имеет, не должен иметь ее и трехмерный ее аналог. Простейший способ определить трехмерную сферу состоит в том, чтобы в точности воспроизвести координатную геометрию обычной сферы. При этом возникает пространство, которое довольно трудно зрительно представить: я не могу показать вам модель в трех измерениях, поскольку трехмерная сфера хотя и имеет всего три измерения, не вкладывается в обычное трехмерное пространство. Для нее необходимо четырехмерное пространство.

Традиционная единичная сфера в трехмерном пространстве включает в себя все точки, расположенные на расстоянии 1 от заданной точки — центра сферы. Аналогично, единичная трехмерная сфера в четырехмерном пространстве включает в себя все точки, расположенные на единичном расстоянии от ее центра. В системе координат мы можем записать формулу для этих точек, воспользовавшись для определения расстояния обобщением теоремы Пифагора[34]. В более общем случае трехмерная сфера представляет собой *любое* пространство, топологически эквивалентное единичной трехмерной сфере, в точности так же, как всевозможные выпуклые версии единичной двумерной сферы топологически являются двумерными сферами. Разумеется, то же относится и к более высоким размерностям.

Если этого вам недостаточно и нужен более геометрический образ, попробуйте вот что: трехмерную сферу можно представить как заполненный шар, вся поверхность которого отождествляется с точкой. Это еще один пример применения правила склеивания. В данном случае процесс аналогичен одному из способов превращения круглого диска в двумерную сферу. Если протянуть нитку вдоль края тканевого диска, а затем туго стянуть ее, как будто затягивая торбу, то результат будет топологически идентичен двумерной сфере. А теперь проведите аналогичную операцию с шаром, но не пытайтесь зрительно представить себе результат: просто представьте шар и как бы приложите к нему правила склеивания.

В любом случае Пуанкаре очень интересовался трехмерной сферой, потому что это, предположительно, простейшее трехмерное топологическое пространство конечной протяженности, не имеющее границы. В 1900 г. он опубликовал статью, в которой объявил, что группы гомологий представляют собой достаточно мощный инвариант, чтобы топологически охарактеризовать трехмерную сферу. А именно, если трехмерное топологическое пространство обладает теми же группами гомологий, что и трехмерная сфера, то оно топологически эквивалентно трехмерной сфере (т. е. может непрерывно в нее преобразовываться). К 1904 г., однако, он обнаружил, что это заявление ошибочно. Существует по крайней мере одно трехмерное пространство, которое не является трехмерной сферой, но имеет те же группы гомологий, что и она. Это пространство стало настоящим триумфом подхода, связанного с правилами склеивания, а доказательство того, что это не трехмерная сфера, привело к созданию нового инварианта, заведомо более мощного, чем гомология.

Сначала о пространстве. Оно известно как додекаэдрическое пространство Пуанкаре, потому что в современном построении используется именно заполненный додекаэдр. Пуанкаре не подозревал о родстве своего пространства с додекаэдром. Сам он поступил иначе: склеил два заполненных тора весьма неочевидным способом. Додекаэдрическую интерпретацию опубликовали в 1933 г., через 21 год после смерти Пуанкаре, Герберт Зейферт и Константин Вебер, и она намного проще для понимания. Аналогия, которую здесь следует помнить, это получение тора путем склеивания противоположных сторон квадрата. Как всегда, не нужно пытаться *действительно* что-то склеить, — достаточно просто помнить, что соответствующие точки рассматриваются именно таким образом. Теперь мы проведем ту же операцию, но возьмем для этого противоположные грани додекаэдра (см. рис. 38).

Пифагорейцы знали о додекаэдре еще 2500 лет назад. Граница додекаэдра состоит из 12 правильных пятиугольников, соединенных в приблизительно сферическую решетку. В каждой его вершине встречаются три пятиугольника. А теперь склеим каждую грань с противоположной… Только для этого

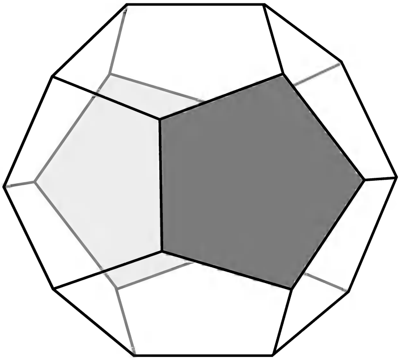


Рис. 38. Чтобы получить додекаэдрическое пространство Пуанкаре, возьмите додекаэдр и склейте попарно все противоположные его грани (такие, как затененные грани на рисунке) с разворотом, чтобы они полностью совпали

их нужно перекрутить. Буквально. Каждую грань, чтобы она совпала с противоположной, нужно повернуть на подходящий угол. Угол берем наименьший из тех, что позволяют совместить соответствующие грани, т. е. 36°. Можно считать это правило своеобразной версией правила изготовления ленты Мёбиуса: конец ленты нужно повернуть на 180°, а затем склеить с противоположным.

Так, пространство получено. А теперь посмотрим на инвариант. Нет, я не растекаюсь мыслью по древу: все это нам потребуется для понимания гипотезы Пуанкаре.

Пуанкаре назвал свой новый инвариант фундаментальной группой. Мы до сих пор пользуемся этим термином, но иногда называем его и иначе: первой гомотопической группой. Гомотопия — это геометрическая конструкция, которая целиком размещается внутри пространства и несет в себе информацию о топологическом типе этого пространства. Она делает это при помощи абстрактной алгебраической структуры, известной как группа. Группа — это набор математических объектов, таких, что комбинация любых двух подобных объектов дает еще один объект той же группы. Для закона комбинирования — его часто называют сложением или умножением, даже если это не те простые операции, которые мы знаем из арифметики — должны выполняться несколько простых и естественных условий. Если мы называем операцию сложением, основные условия такие:

- группа содержит элемент, который ведет себя как нуль: при добавлении к любому другому элементу группы ничего не меняется;
- каждый элемент имеет в группе соответствующий ему элемент с противоположным знаком: при сложении такой пары получается нуль;
- при сложении трех элементов группы не имеет значения, какие два вы складываете первыми. Иными словами, $(a + b) + c = a + (b + c)$. Это называется законом ассоциативности.

Единственный алгебраический закон, который не считается обязательным (хотя иногда и выполняется), — это закон коммутативности[35] $a + b = b + a$.

Фундаментальная группа Пуанкаре представляет собой своего рода упрощенный скелет пространства. Это топологический инвариант: топологически эквивалентные пространства имеют одну и ту же фундаментальную группу. Чтобы лучше разобраться в этом полезном понятии и, очень может быть, отчасти восстановить мотивы Пуанкаре, посмотрим, как это работает, на примере окружности. Воспользуемся образом, который восходит еще к Гауссу: представьте себе муравья, вся вселенная которого ограничена окружностью. Как он может определить, какой формы его вселенная? Сумеет ли он отличить окружность от, скажем, прямой линии? Не забывайте, что муравей не может выйти за пределы своей вселенной, не может взглянуть на нее со стороны и понять, что она круглая. Он может лишь бродить по вселенной, что бы она собой ни представляла. В частности, муравей не в состоянии понять, что его вселенная изогнута, потому что и свет в ней движется только по кругу. И не обращайте, пожалуйста, внимания на практические сложности, к примеру, на то, что объектам придется, встречаясь, проходить сквозь друг друга, — в любом случае наша аналогия достаточно свободна.

Муравей может определить форму вселенной несколькими способами. Я сосредоточусь на методе, который можно обобщить на любые топологические пространства. Для целей данного обсуждения муравей — точка. Он живет на автобусной остановке, которая тоже представляет собой точку. Каждый день муравей выходит из домика, садится в автобус (который, конечно, тоже точка), а вечером возвращается обратно. Самый простой маршрут — № 0: он просто стоит на остановке и никуда не едет. Для более интересной экскурсии муравей садится в автобус № 1, который объезжает вселенную ровно один раз против часовой стрелки и останавливается, вернувшись домой. Автобус № 2 объезжает вселенную дважды, № 3 — трижды и т. д.; один автобус, движущийся против часовой стрелки, для каждого положительного целого числа. Есть и отрицательные автобусы, которые ездят в противоположном направлении. Автобус № – 1 объезжает вселенную один раз по часовой стрелке, № – 2 — два раза и т. д.

Муравей быстро замечает, что две последовательные поездки на автобусе № 1, по существу, эквивалентны одной поездке на № 2, а три поездки на № 1 — одной поездке на № 3. Аналогично, следующие одна за другой поездки на автобусах № 5 и № 8 соответствуют одной поездке на автобусе № 13. Более того, для любых двух положительных номеров поездка на автобусе с первым номером плюс следующая за ней поездка на автобусе со вторым номером сводится к поездке на автобусе с номером, соответствующим их сумме.

Следующий шаг тоньше. *Примерно* то же соотношение сохраняется для автобусов с отрицательными номерами и для № 0. Поездка на № 0 плюс поездка на № 1 очень похожа на поездку на № 1. Однако есть и небольшая разница. В поездке

0 + 1 автобус № 0 некоторое время стоит на остановке, отрабатывая свой маршрут, а в поездке только на № 1 ничего подобного не происходит. Поэтому мы вводим понятие со странным названием гомотопия («то же место» по-гречески). Две петли гомотопичны, если одна из них может быть непрерывно преобразована в другую. Если мы позволим гомотопиям менять расписание автобусов, можно будет постепенно снизить время, которое муравей проводит в стоящем на остановке автобусе № 0, и, в конце концов, период сидения на месте просто исчезнет. Теперь между поездкой 0 + 1 и поездкой 1 нет никакой разницы, так что «с точностью до гомотопии» результат — это просто поездка на автобусе № 1. Иными словами, уравнение для автобусных номеров 0 + 1 = 1 остается верным не для поездок, а для гомотопических классов поездок.

А если за поездкой на автобусе № 1 последует поездка на автобусе № – 1? Нам хотелось бы, чтобы в ответе стояла поездка № 0, но это не так. Автобус в этом случае проезжает весь путь сначала против часовой стрелки, а потом — обратно. Это далеко не то же самое, что провести все время поездки в стоящем на остановке автобусе. Поэтому 1 + (–1), т. е. 1 – 1, не равно 0. На помощь опять же приходит гомотопия. Комбинация автобусов 1 и –1 в целом гомотопна поездке на автобусе 0. Чтобы понять, почему, представьте, что муравей следует по суммарному маршруту автобусов 1 и –1 на автомобиле, но, чуть-чуть не доехав до остановки, разворачивается и едет назад. Такая поездка очень близка к двойной поездке на автобусе: пропущен всего лишь крохотный кусочек маршрута. Таким образом, первоначальное двойное путешествие непрерывно уменьшилось и превратилось в немного более короткую поездку на машине. Теперь муравей может снова чуть-чуть укоротить поездку, повернув назад чуть раньше. Он может таким образом укорачивать поездку, разворачивая автомобиль все раньше и раньше, пока не окажется просто сидящим на остановке. Процесс сжимания поездки — тоже гомотопия. Она показывает, что поездка 1 плюс поездка –1 гомотопна поездке на автобусе № 0. Иными словами, 1 + (–1) = 0 для гомотопических классов поездок.

Теперь любой алгебраист без труда сможет доказать, что поездка на автобусе любого маршрута плюс вторая поездка на каком-нибудь автобусе гомотопна поездке на автобусе, номер которого получается сложением двух автобусных номеров. Это верно для положительных автобусов, для отрицательных автобусов и для автобуса № 0. Так что если мы складываем поездки — или, вернее, гомотопические классы поездок, — то получаем группу. Более того, очень знакомую группу. Ее элементами являются целые числа (номера автобусов), а ее операцией — сложение. Такая группа традиционно обозначается символом Z от немецкого слова Zahl (“целый”).

Гораздо труднее, но все же можно доказать, что в кольцевой вселенной *любая* кольцевая автомобильная поездка — даже если она предусматривает множество возвратов, отступлений или метаний взад-вперед на одном и том же участке дороги — гомотопична одной из стандартных автобусных поездок. Более того, автобусные поездки с разными номерами не гомотопичны. Доказательство требует некоторых теоретических познаний. Его основа — гауссов порядок кривой, или число вращения. Это число полных обходов окружности против часовой стрелки, которое совершает муравей за всю поездку[36], и это номер маршрута, которому гомотопична ваша конкретная поездка.

Если заполнить все пробелы и расставить все точки над i, это описание доказывает, что фундаментальная группа окружности совпадает с группой целых чисел Z по операции сложения. Чтобы складывать поездки, нужно просто складывать соответствующие им числа вращения. При помощи этого топологического инварианта муравей может отличить свою кольцевую вселенную от, скажем, бесконечной прямой линии. На прямой любая поездка, как ни мечись, в какой-то момент должна достичь максимально удаленной от дома точки. Тогда мы можем непрерывно сжать поездку, постепенно уменьшая все расстояния от дома в одной и той же пропорции — сначала до 99%, затем до 98% и т.д. Поэтому на прямой *любая* поездка гомотопна нулю: можно просто остаться дома. Фундаментальная группа прямой содержит только один элемент: 0. Ее алгебраические свойства тривиальны: 0 + 0 = 0, и называется она тривиальной группой. А поскольку тривиальная группа не совпадает с группой целых чисел, муравей может понять, живет ли он на прямой или на окружности.

Как я уже говорил, существуют и другие методы, но именно так муравей может заметить разницу при помощи фундаментальной группы Пуанкаре.

А теперь предположим, что наш муравей живет на поверхности и это опять же вся его вселенная. Он не может отойти в сторону и посмотреть, какая именно поверхность является его домом. Может ли он разобраться в топологии своей вселенной? В частности, сможет ли он различить сферу и тор? Ответ по-прежнему «да», а метод тот же, при помощи которого мы исследовали вселенную-окружность: сесть в автобус и совершать круговые поездки, которые начинаются и заканчиваются в одной точке — дома. Чтобы сложить такие поездки, их нужно проделать по очереди — одну за другой. Нулевая поездка — это остаться дома; поездка с обратным знаком — это точно такая же поездка в противоположном направлении. Работая с гомотопическими классами поездок, мы получим группу. Это фундаментальная группа поверхности. По сравнению с вселенной-окружностью здесь куда больше свободы в выборе маршрутов поездок и непрерывном преобразовании их в другие поездки; тем не менее основная идея та же.

Фундаментальная группа здесь тоже является топологическим инвариантом, и муравей может воспользоваться ею, чтобы выяснить, живет ли он на сфере или на торе. Если его вселенная — сфера, то любая поездка, совершенная муравьем, может быть постепенно преобразована в нулевую поездку — пребывание дома. Однако в случае, если вселенная — тор, это не так. Некоторые поездки могут быть преобразованы в нуль, но с поездкой, которая хотя бы раз обойдет вокруг центрального отверстия (см. рис. 39 слева), ничего подобного проделать нельзя. Это утверждение нуждается в доказательстве, но это не проблема. На торе тоже есть стандартные поездки, но теперь номера автобусов представляют собой пары целых чисел (m, n). Первое число m указывает, сколько раз маршрут проходит сквозь центральное отверстие. Второе число n указывает, сколько раз маршрут обвивается вокруг тора. На рис. 39 справа показан маршрут (5, 2), который пять раз проходит сквозь отверстие и дважды обвивается вокруг тора. Чтобы сложить поездки, нужно сложить соответствующие числа маршрутов, к примеру: $(3, 6) + (2, 4) = (5, 10)$. Фундаментальная группа тора — группа *пар* целых чисел.

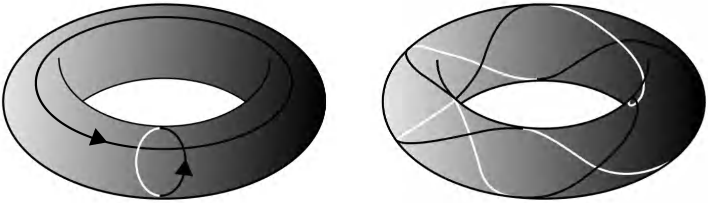


Рис. 39. Автобусные поездки (1, 0) и (0, 1) на торе *(слева)*. Поездка (5, 2) *(справа)*. Серым цветом показаны невидимые участки линий

Любое топологическое пространство имеет фундаментальную группу, определенную в точности так же, с использованием поездок — или, точнее, петель, — которые начинаются и заканчиваются в одной точке. Пуанкаре придумал фундаментальную группу, чтобы доказать, что его додекаэдрическое пространство не является трехмерной сферой, хотя и имеет те же гомологические инварианты. Его первоначальный метод прекрасно приспособлен к вычислению фундаментальной группы. Более современный метод «скручивания и склеивания» приспособлен к нему еще лучше. Ответом оказывается группа из 120 элементов, связанная с додекаэдром. А вот фундаментальная группа трехмерной сферы, напротив, состоит лишь из одного элемента: нулевой петли. Так что додекаэдрическое пространство топологически не эквивалентно сфере, несмотря на одинаковые группы гомологий, и Пуанкаре доказал, что утверждение, сделанное им в 1900 г., ошибочно.

Пуанкаре продолжал рассуждать о своем новом инварианте: может быть, это и есть недостающий ингредиент топологической характеристики трехмерной сферы? А может, любое трехмерное пространство с той же фундаментальной группой, как у трехмерной сферы, т. е. с тривиальной группой, должно *на самом деле* быть трехмерной сферой? Он сформулировал это предположение как отрицание в виде вопроса: «Рассмотрим компактное трехмерное многообразие [топологическое пространство] V, не имеющее границы. Возможно ли, чтобы фундаментальная группа многообразия V была тривиальной, хотя V не есть трехмерная сфера [топологически не эквивалентно ей]?» Он оставил вопрос открытым, но очень правдоподобно мнение, что ответ при такой постановке вопроса очевиден — «нет». И вскоре это предположение получило известность как гипотеза Пуанкаре. И столь же быстро стало одним из самых знаменитых открытых вопросов топологии.

Фраза «тривиальная фундаментальная группа» означает, в сущности, что «любая петля может быть непрерывно преобразована в точку». Таким свойством обладает не только трехмерная сфера, но и любая аналогичная ей n-мерная сфера любой размерности n. Так что мы можем выдвинуть точно такое же предположение для сферы любой размерности. Такое утверждение известно как n-мерная гипотеза Пуанкаре. Это верно для $n = 2$, согласно теореме классификации для поверхностей. И это все, чего удалось достичь математикам за 50 с лишним лет.

В 1961 г. Стивен Смейл взял прием классификации поверхностей и применил его к более высоким измерениям. Один из способов представить себе тор с g отверстиями заключается в том, чтобы взять сферу и приделать к ней мысленно g ручек —

точно таких, какие бывают у чайной чашки или кружки. Смейл обобщил это построение для любой размерности и назвал процесс разложением на ручки. Он проанализировал, как могут изменяться ручки при неизменной топологии пространства, и вывел гипотезу Пуанкаре во всех размерностях, больших или равных 7. Для более низких размерностей его доказательство не работало, но другие математики нашли способы с этим справиться: Джон Столлингс провел доказательство для размерности 6, а Кристофер Зиман — для 5. Однако один из существенных этапов доказательства, известный как трюк Уитни, упрямо отказывался работать в размерностях 3 и 4, потому что в таких пространствах просто не хватает места для необходимых маневров, и никто не мог найти эффективной замены этому приему. Постепенно сформировалось мнение о том, что топология пространств для этих двух размерностей может оказаться весьма необычной.

Это мнение, однако, было поколеблено в 1982 г., когда Майкл Фридман получил доказательство четырехмерной гипотезы Пуанкаре, для которого не требовался трюк Уитни. Доказательство было чрезвычайно сложным, но работало. Итак, после 50 лет топтания на месте и 20 лет лихорадочной активности топологи расправились наконец с гипотезой Пуанкаре для всех размерностей, кроме той, о которой, собственно, и шла речь изначально. Успехи впечатляли, но методы, при помощи которых они были достигнуты, не позволяли сказать почти ничего о трехмерном случае. Требовался новый подход.

Перечень того, что позволило, наконец, сдвинуться с мертвой точки, отчасти напоминает традиционный список подарков к свадьбе: что-то старинное, антикварное, что-то новенькое, что-то взятое взаймы и, наконец, если немного выходить за рамки, что-то из даров небес. Старинная идея заключалась в обращении к той области топологии, которая на фоне активной работы с пространствами более высоких размерностей представлялась почти исчерпанной: в топологию поверхностей.

Новая идея была в том, чтобы заново рассмотреть классификацию поверхностей с позиции, на первый взгляд, совершенно чуждой: с позиции классической геометрии. Одолженной идеей можно считать поток Риччи, источником вдохновения для которого послужил математический аппарат общей теории относительности Эйнштейна. Ну а к дарам небес можно отнести нечто вроде попадания «пальцем в небо»: далеко идущие предположения, опирающиеся отчасти на интуицию, но куда больше — на надежду.

Вспомним, что ориентируемые поверхности без границы можно проклассифицировать: каждая из них топологически эквивалентна тору с некоторым числом отверстий. Это число — род поверхности, и когда род равен нулю, поверхность представляет собой сферу без ручек, т. е. просто сферу. Это слово сразу же напоминает нам о том, что среди всех топологических сфер одна поверхность стоит особняком и является архетипом. Конкретно речь идет о единичной сфере в евклидовом пространстве. Забудьте на мгновение все разговоры о резиновом листе — пока отложим это в сторону. Сосредоточьтесь на старой доброй евклидовой сфере. У нее много разных дополнительных математических свойств, проистекающих из жесткости и однозначности евклидовой геометрии. Важнейшее из этих свойств — кривизна. Кривизну можно квантифицировать: для каждой точки геометрической поверхности существует число, говорящее о том, насколько изогнута поверхность вблизи этой точки. Сфера — единственная в евклидовом пространстве замкнутая поверхность, кривизна которой во всех точках одинакова и положительна.

Это странно, потому что постоянная кривизна — не топологическое свойство. Еще загадочнее то, что сфера не одинока. Существует еще одна стандартная геометрическая поверхность, которая стоит особняком и представляет собой архетипический тор. А именно: начнем с квадрата на плоскости и отождествим противоположные его стороны (см. рис. 12 из главы 4). Результат в трехмерном пространстве после скатывания рулона и соединения тождественных сторон выглядит изогнутым. Однако, по существу, мы можем работать непосредственно с квадратом, применив дополнительно правила склеивания. Квадрат имеет естественную геометрическую структуру: это участок на евклидовой плоскости. Плоскость, кстати говоря, тоже имеет постоянную кривизну, на этот раз *нулевую*. Тор с данной конкретной геометрией тоже имеет нулевую кривизну и называется *плоским тором*. Возможно, название звучит как оксюморон, но для муравья, живущего на плоском торе и пользующегося линейкой и транспортиром для измерения расстояний и углов, местная геометрия вполне соответствовала бы плоской геометрии.

Геометры XVIII в., стараясь разобраться в аксиоме Евклида о существовании параллельных линий, пытались вывести ее из остальных евклидовых постулатов, но раз за разом терпели поражение. В конце концов пришло понимание, что такой вывод невозможен. Существует три различных типа геометрии, в каждом из которых выполняются все условия и требования Евклида, за исключением аксиомы о параллельных прямых. В настоящее время эти геометрии известны как евклидова (это плоскость, на которой аксиома о параллельных прямых верна), эллиптическая (геометрия на поверхности сферы с некоторыми финтифлюшками: здесь две прямые всегда пересекаются, а параллельной прямой не существует) и гиперболическая геометрия (где некоторые прямые не пересекаются, а параллельная прямая не единственна). Более того, классические математики интерпретируют эти геометрии как геометрии искривленных пространств. Евклидова геометрия соответствует нулевой кривизне, эллиптическая/сферическая геометрия — постоянной положительной кривизне, а гиперболическая геометрия — постоянной отрицательной кривизне.

Мы только что видели, как можно получить две из трех перечисленных геометрий: они возникают на сфере и на плоском торе. В терминах теоремы классификации это торы рода g для $g = 0$ и 1. Единственное, чего у нас пока не хватает, это

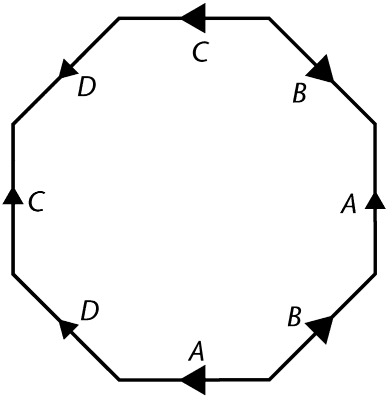


Рис. 40. Изготовление тора с двумя отверстиями из восьмиугольника путем попарного отождествления сторон (*AA*, *BB*, *CC*, *DD*)

гиперболической геометрии. Может быть, каждый тор с g дырками обладает естественной геометрической структурой, основанной на том, что в гиперболическом пространстве взяли некий многоугольник и отождествили у него некоторые стороны? Ответ поразителен: «да» для *любой* величины g, большей или равной 2. На рис. 40 показан пример для g = 2 на основе восьмиугольника. Я опущу гиперболическую геометрию и идентификацию этой поверхности как двумерного тора, но скажу, что разобраться в этом можно. Различные g возникают, если мы берем разные многоугольники, но исключений нет — можно получить любой g. Используя профессиональную лексику, скажем, что тор с двумя и более дырками имеет естественную гиперболическую структуру. Теперь можно пересмотреть список стандартных поверхностей:

- сфера: g = 0 — эллиптическая геометрия;
- тор: g = 1 — евклидова геометрия;
- тор с g дырками: g = 2, 3, 4... — гиперболическая геометрия.

Может показаться, что мы выплеснули с водой и ребенка, ведь топология должна иметь дело с геометрией на резиновом листе, а не с жесткой геометрией. Но теперь мы легко можем вернуть резину на место. Жесткая геометрия используется здесь только для того, чтобы *определить* стандартные поверхности. Она позволяет дать простые описания, которые оказываются еще более жесткими. А теперь ослабим жесткость, т. е. позволим пространству *стать* резиновым и разрешим деформироваться, что невозможно при жесткой структуре. При этом мы получим поверхности, топологически эквивалентные стандартным, но не получаемые из них путем жестких сдвигов. Согласно теореме о классификации, таким образом можно получить любую топологическую поверхность.

Топологи знали о существовании такой связи между геометрией и теоремой о классификации поверхностей, но в то время она представлялась забавным совпадением, дающим, несомненно, весьма ограниченные возможности в двух измерениях. Все понимали, что трехмерный случай намного богаче и, в частности, пространствами постоянной кривизны его возможности не исчерпываются. Но понять, что жесткая геометрия может оказаться полезной при рассмотрении трехмерной топологии, сумел лишь Уильям Терстон — один из лучших геометров мира. Несколько указаний на это уже имелось: трехмерная сфера Пуанкаре, исходя из ее определения, обладает естественной эллиптической/сферической геометрией. Хотя стандартный додекаэдр обитает в евклидовом пространстве, угол между его смежными гранями меньше 120°, так что три таких угла не образуют полной окружности. Чтобы исправить это, нам придется слегка надуть додекаэдр, чтобы его грани стали немного выпуклыми: это сразу превращает естественную геометрию фигуры из евклидовой в сферическую. Аналогично, треугольники на сфере тоже становятся выпуклыми. Трехмерный тор, полученный путем отождествления противоположных граней куба, обладает плоской, т. е. евклидовой, геометрией, в точности так же, как его двумерный аналог. Макс Ден и другие исследователи открыли несколько трехмерных топологических пространств, обладающих естественной гиперболической геометрией.

У Терстона появились первые подозрения о возможности существования общей теории, но, чтобы она обрела хотя бы относительную правдоподобность, требовались два нововведения. Во-первых, необходимо было расширить диапазон трехмерных геометрий. Исходя из здравого смысла, Терстон сформулировал некоторые условия и выяснил, что им удовлетворяет ровным счетом восемь геометрий. Три из них — это классика: сферическая, евклидова и гиперболическая геометрия. Еще две напоминают цилиндры: плоские в одном направлении, изогнутые в двух других. Изогнутая часть имеет либо положительную кривизну, как у двумерной сферы, либо отрицательную, как у гиперболической плоскости. Наконец, есть еще три, достаточно формальные, геометрии.

Во-вторых, было ясно, что некоторые трехмерные пространства не поддерживают ни одну из восьми геометрий. Но нашелся и выход: разрезать пространство на куски. Один кусок, возможно, обладает сферической геометрической структурой, другой — гиперболической и т. д. Чтобы разрезание было полезным, его надо проводить по очень строгим правилам, чтобы обратный процесс — собирание кусков в единое целое — позволил получить полезную информацию. Хорошей новостью стало то, что во многих случаях это возможно. В 1982 г. Терстон в приступе вдохновения сформулировал гипотезу о геометризации: любое трехмерное пространство может быть разрезано на куски, каждый из которых обладает естественной геометрической структурой, соответствующей одной из восьми возможных геометрий. Он доказал также, что если его гипотеза о геометризации верна, то гипотеза Пуанкаре окажется простым ее следствием.

Тем временем появилось и второе направление атаки, тоже геометрическое и тоже основанное на кривизне, но исходящее из совершенно иной области: математической физики. Гаусс, Риман и целая школа итальянских геометров создали общую теорию искривленных пространств, получивших название многообразий, причем концепция расстояния у них необыкновенно расширила и евклидову, и классическую неевклидову геометрию. Кривизна уже не обязана быть постоянной: она может плавно меняться от одного конца к другому. К примеру, фигура, напоминающая собачью косточку, имеет положительную кривизну на концах, но отрицательную посередине, и величина кривизны изменяется плавно от одного участка к другому. Кривизна квантифицируется при помощи математических инструментов, известных как тензоры. Около 1915 г. Альберт Эйнштейн понял, что тензоры кривизны — это именно то, чего ему не хватало для расширения специальной теории относительности, описывающей пространственно-временные отношения, до общей теории относительности, включающей также и гравитацию. В этой теории гравитационное поле представлено как кривизна пространства, а эйнштейновы уравнения поля описывают, как соответствующая мера кривизны — тензор кривизны — изменяется в зависимости от распределения материи. В результате кривизна пространства *плывет* со временем; вселенная или некая ее часть спонтанно меняет форму.

Ричард Гамильтон, специалист по римановой геометрии, понял, что тот же фокус можно применить в более общем плане и что результатом этого может стать доказательство гипотезы Пуанкаре. Идея состояла в том, чтобы работать с одной из простейших мер кривизны, именуемой кривизной Риччи в честь итальянского геометра Грегорио Риччи-Курбастро. Гамильтон записал уравнение, определявшее, как кривизна Риччи должна изменяться со временем: уравнение потока Риччи. Согласно этому уравнению, кривизна должна была постепенно перераспределиться и стать как можно более равномерной. Картина немного напоминает кошку под ковром из главы 4, но теперь кошка, хотя и не может сбежать, способна растечься по полу ровным слоем. (Говоря иначе, кошка здесь должна быть топологической.)

К примеру, в двумерном случае начнем с грушевидной поверхности (см. рис. 41). На одном конце она имеет область сильной положительной кривизны. Область на другом, более толстом конце тоже положительно искривлена, но не так сильно, а в промежутке грушу опоясывает область с отрицательной кривизной. По существу, поток Риччи переносит кривизну с сильно искривленного конца (и в меньшей степени с другого конца) в отрицательно искривленную область до тех пор, пока вся отрицательная кривизна не будет поглощена. На этой стадии результат — бугристая поверхность с повсеместно положительной кривизной. Поток Риччи продолжает перераспределять кривизну, забирая ее из сильно искривленных областей и перенося в менее искривленные. Время идет, и поверхность становится все ближе и ближе к той единственной форме, что имеет постоянную положительную кривизну, т. е. к евклидовой сфере. Топология остается прежней, хотя форма, если посмотреть подробнее, меняется. Следуя потоку Риччи, можно доказать что первоначальная грушевидная поверхность топологически эквивалентна сфере.

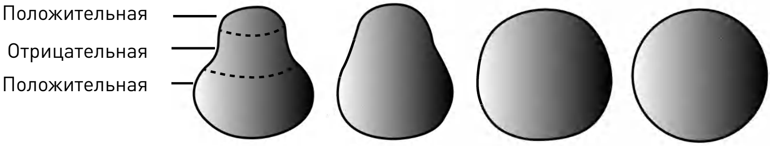


Рис. 41. Как поток Риччи превращает грушу в шар

В этом примере топологический тип поверхности был очевиден с самого начала, однако та же общая стратегия действует для любого многообразия. Начните со сложной формы и следуйте за потоком Риччи. Со временем кривизна перераспределяется более равномерно, и форма упрощается. В конце концов вы должны получить простейшую форму с той же топологией, что и у первоначального многообразия, какой бы эта топология ни была. В 1981 г. Гамильтон доказал, что такая стратегия работает в двух измерениях, обеспечивая новое доказательство теоремы о классификации для поверхностей.

Кроме того, он добился значительного прогресса в аналогичной стратегии для трехмерных многообразий, но здесь возникло серьезное препятствие. В двух измерениях любая поверхность автоматически упрощается, следуя потоку Риччи. Это верно и в трех измерениях, если первоначальное многообразие во всех точках имеет строго положительную кривизну и нигде — нулевую или отрицательную. К несчастью, если в многообразии есть точки с нулевой кривизной — а они часто есть, — пространство, двигаясь в потоке, может запутаться. При этом возникают сингулярности — места, где многообразие перестает быть гладким. В таких точках уравнение потока Риччи не работает, и перераспределение кривизны прекращается. Естественный способ обойти это препятствие заключается в том, чтобы понять, что представляют собой сингулярности, и изменить многообразие — может быть, разрезать его на куски, чтобы можно было дать стартовый толчок потоку Риччи. Такая стратегия может оказаться успешной, если вы в достаточной степени контролируете связь топологии измененного многообразия к первоначальной. К несчастью, Гамильтон понял также, что для трехмерных пространств сингулярности потока Риччи могут быть чрезвычайно сложными — судя по всему, слишком сложными, чтобы применять подобные уловки. В общем, поток Риччи быстро стал в геометрии стандартным методом, но для доказательства гипотезы Пуанкаре его не хватило.

К 2000 г. гипотеза по-прежнему оставалась не доказанной; после вхождения в число семи проблем тысячелетия она приобрела еще более широкую известность и признание. К тому моменту стало ясно, что если каким-то образом удастся все же добиться, чтобы идея Гамильтона сработала, то тем самым будет доказана не только гипотеза Пуанкаре, но и гипотеза Терстона о геометризации. Приз был соблазнителен и близок, но в руки не давался.

В математике, как и в остальных отраслях науки, работа, чтобы ее признали, должна быть опубликована, а для этого —

пройти рецензирование. Специалисты в соответствующей области должны внимательно прочитать работу, проверить логические выкладки и убедиться в безошибочности вычислений. Для сложной и значительной математической работы этот процесс может занять немало времени. Как упоминалось в главе 4, раньше выходом в каких-то ситуациях становился препринт, но сегодня существует стандартный веб-сайт arXiv.org, своеобразный архив, где после частичного рассмотрения и утверждения (чтобы отсечь всякие глупости) разрешается размещать электронные препринты. В настоящее время большинство исследователей знакомится с новыми результатами на сайте arXiv или на собственном сайте автора.

В 2002 г. Григорий Перельман разместил на сайте arXiv препринт о потоке Риччи. В работе было сделано замечательное утверждение: поток Риччи градиентоподобен. Иными словами, существует вполне определенное направление вниз — единственная числовая величина, связанная с формой многообразия, и многообразие всегда течет вниз в том смысле, что эта величина всегда уменьшается со временем. Она чем-то напоминает высоту в ландшафте и позволяет количественно оценить «упрощение» многообразия. Градиентоподобные потоки имеют немало ограничений: к примеру, они не могут ходить кругами или вести себя хаотично. Никто, похоже, не подозревал, что поток Риччи окажется таким ручным. Но Перельман не просто выдвинул предположение: он доказал это. В конце он наметил цепочку рассуждений, которые должны были бы доказать гипотезу Терстона о геометризации — а она, если помните, подразумевает гипотезу Пуанкаре, но заходит на самом деле гораздо дальше, — и пообещал подробнее изложить все это в следующих статьях на сайте arXiv. В течение следующих восьми месяцев он разместил там две статьи, содержавшие большую часть обещанных подробностей.

Первая статья вызвала немалый переполох. Перельман утверждал, что ему удалось реализовать всю программу Гамильтона — использовать поток Риччи для упрощения трехмерного многообразия и доказать, что результат получился в точности таким, как предсказывал Терстон. Две другие статьи добавили рассуждениям Перельмана убедительности: у математиков возникло чувство, что это человек знает, о чем говорит, и что его идеи — не просто очередная правдоподобная стратегия с неизменной логической прорехой или недоказанным допущением. Обычный скепсис математического сообщества по отношению к любым заявлениям о решении одной из великих задач слегка поутих. Возникло ощущение, что его попытка вполне может увенчаться успехом.

Однако дьявол, как всегда, кроется в деталях, а в математике детали бывают дьявольски непокорными! Работу необходимо было проверить, не спеша и на полную глубину, причем сделать это должны были те, кто разбирается в соответствующих областях и в состоянии распознать потенциальные ловушки. А это было непросто, поскольку Перельман в своей работе свел воедино по крайней мере четыре очень разные области математики и математической физики, а мало кто из математиков может похвастать знаниями более чем в одной-двух областях. Анализ корректности его доказательства потребовал бы много усилий и командной работы. Более того, в препринтах на сайте arXiv не было всех подробностей, необходимых в публикуемой статье. Для препринтов они были написаны довольно ясно, но точки над i там были расставлены не все. Так что экспертам нужно было реконструировать часть рассуждений Перельмана — при том, что сам-то он занимался этой работой несколько лет!

На все это требовалось время. Перельман читал лекции по своему доказательству и отвечал по электронной почте на вопросы, касавшиеся различных его этапов. Всякий раз, как кто-нибудь находил кажущуюся прореху, он быстро откликался, объяснял необходимое и заполнял пробелы. Все выглядело обнадеживающе. Однако никто не собирался рисковать репутацией и заявлять публично, что Перельман доказал гипотезу Пуанкаре и, тем паче, еще более сложную гипотезу о геометризации. Нужна была полная уверенность в том, что доказательство безошибочно. Поэтому, несмотря на общее благосклонное отношение к работе Перельмана, публичного признания она поначалу не получила. Это было ожидаемо, но время шло, и Перельмана все больше охватывало раздражение, потому что, как ему казалось, он впустую тратил время. Он-то *знал,* что его доказательство верно, и не мог понять, почему у остальных возникают такие сложности. Он отказался написать о своей работе подробнее или представить статью в какой-нибудь журнал. С его точки зрения, дело было сделано, а препринты на arXiv содержали всю необходимую информацию. Он перестал отвечать на вопросы, касавшиеся недостающих вроде бы деталей. Для него все было очевидно. Да ладно, ребята, вы и сами можете разобраться в этом, без моей помощи. Это не так уж сложно.

Некоторые писали, что математическое сообщество было несправедливо к Перельману. Но те, кто так говорят, просто не понимают, как принято действовать, когда появляется заявка на решение одной из великих задач. Было бы безответственно просто похлопать автора по плечу, сказать: «Отлично! Молодец!» — и забыть о том, чего не хватает в его препринтах. Вполне справедливо было попросить его подготовить более подробное изложение доказательства, пригодное для публикации. Когда речь идет о столь важной задаче, спешить нельзя. Специалисты из кожи вон лезли, тратили кучу времени на доказательство Перельмана и больше обычного старались сдержать свой естественный скептицизм. Сказать по правде, к автору отнеслись даже *более* благожелательно, чем обычно. И со временем, когда процесс проверки был завершен, его работу приняли и признали.

К этому моменту, однако, Перельман успел потерять терпение. Возможно, сказалось и то, что решенная им задача была настолько значительной, что ничто, по существу, уже не могло с ней сравниться. Он был как альпинист, сумевший подняться на Эверест в одиночку и без кислорода. Сравнимых вызовов просто не осталось. Успех в средствах массовой информации его не прельщал: он ждал признания со стороны равных, а не со стороны телеведущих всех сортов. Потому можно понять, почему, когда коллеги наконец признали, что он прав и предложили ему Филдсовскую медаль и премию Института Клэя, он не захотел принять эти награды.

Доказательство Перельмана отличается глубиной и элегантностью и открывает перед исследователями целый новый мир топологии. Автор сумел реализовать план Гамильтона по потоку Риччи, придумав хитрые способы обойти существование сингулярностей. Один из таких способов заключается в том, чтобы изменить масштабы пространства и времени и таким образом избавиться от сингулярности. Когда такой подход не работает, говорят, что сингулярность схлопывается. В подобных случаях Перельман анализирует геометрию потока Риччи в подробностях и разбирает, как именно может произойти схлопывание. По существу, пространство как бы выпускает бесконечно тонкие щупальца, иногда во множестве, как ветви дерева. Если какая-то ветка близка к схлопыванию, ее можно срезать и заменить гладкой крышечкой. Перед некоторыми из этих щупальцев поток Риччи буксует: если так, оставляем их в покое. Если же нет, поток Риччи можно запустить заново. В итоге некоторые щупальца заменяются гладкими крышками, а другие временно прерываются, но поток продолжает работать.

Процедура срезания и замазывания щупалец рубит пространство примерно так же, как терстоново рассечение на куски, каждый со своей геометрией (одной из восьми). Оказывается, что обе процедуры приводят к более или менее одинаковым результатам. Но есть один принципиально важный технический момент: операция обрезки не должна бесконечно ускоряться, так чтобы за конечное время проводилось бесконечное число операций. Это часть доказательства — одна из сложнейших.

Некоторые комментаторы критикуют математическое сообщество за несправедливое отношение к Перельману. Конечно, никто не должен быть закрыт для критики, да и инциденты, в которых, в принципе, можно разглядеть несправедливость или по крайней мере необдуманность, действительно имели место, но в целом математическое сообщество отреагировало на работу Перельмана быстро и положительно. Кроме того, реакция была осторожной, что абсолютно естественно в математике и науке вообще, и не без причин. Неизбежная публичность и слава, еще более яркая благодаря премии в миллион долларов, сказалась бы на любом человеке, и Перельман не исключение.

С момента размещения первой статьи Перельмана на arXiv в ноябре 2002 г. до объявления в марте 2010 г. о присуждении ему премии Института Клэя прошло восемь лет. Кажется, что это серьезная и, возможно, безосновательная задержка. Однако та, первая, публикация содержала лишь часть доказательства. Остальное по большей части было размещено на сайте в марте 2003 г. К сентябрю 2004 г., полтора года спустя после этой второй публикации, сообщество специалистов по потоку Риччи и топологии успело проработать доказательство — следует отметить, что этот процесс начался всего через несколько дней после первой публикации, — и ведущие эксперты объявили, что «поняли его». Они нашли в нем ошибки, нашли пробелы, но выразили уверенность в том, что все это можно исправить. Полтора года — совсем немного, когда речь идет о таком важном вопросе.

В конце 2005 г. Международный математический союз связался с Перельманом и предложил ему Филдсовскую премию, высшую математическую награду. Присудить ее предполагалось на Международном математическом конгрессе в 2006 г. Конгресс проводится раз в четыре года, так что это была бы первая возможность почтить ученого за серьезное достижение. Поскольку в полноте доказательства гипотезы Пуанкаре оставались некоторые сомнения — в нем все еще время от времени обнаруживались ошибки, — премия официально присуждалась за успехи в понимании потока Риччи (эта часть препринтов Перельмана к тому моменту уже считалась свободной от ошибок).

Условия присуждения премии за решение проблем тысячелетия размещены на сайте Института Клэя. В частности, предлагаемое решение должно быть опубликовано в рецензируемом журнале и принято математическим сообществом, причем отношение к нему не должно измениться за два года после публикации. После этого специальный консультативный комитет должен рассмотреть вопрос и выдать рекомендацию: присуждать автору премию или нет. Перельман не выполнил первого условия и, судя по всему, никогда уже этого не сделает. С его точки зрения, препринтов на сайте arXiv достаточно. Тем не менее Институт Клэя махнул на это рукой и объявил о начале уставного двухлетнего срока: требовалось посмотреть, не всплывут ли еще какие-нибудь ошибки или вопросы. Срок истек в 2008 г.; теперь нужно было следовать строгой (чтобы, не дай бог, не выдать премию преждевременно) процедуре.

Это правда, что некоторые эксперты не спешили выражать свою уверенность в корректности доказательства Перельмана. Причина понятна: они действительно не были в ней уверены. Не будет преувеличением сказать, что единственным человеком, способным быстро разобраться в доказательстве Перельмана, мог бы быть только второй Перельман. Невозможно читать математическое доказательство с листа, как музыканты читают ноты. Необходимо убедить самого себя в том, что здесь все разумно и имеет смысл. Всякий раз, когда аргументация усложняется, ты понимаешь, что возрастает и вероятность ошибки. То же можно сказать и о ситуации, когда излагаемые идеи становятся слишком простыми: многие перспективные доказательства споткнулись на утверждениях настолько очевидных, что ничего доказывать, казалось бы, вообще не требовалось. До тех пор, пока эксперты не убедились окончательно в том, что доказательство верно в своей основе — а именно в этот момент они признали достижение Перельмана, *несмотря* на оставшиеся пробелы и ошибки, — разумно было воздержаться от суждений. Вспомните, к примеру, шумиху вокруг холодного синтеза, данные о котором через некоторое время были опровергнуты. Осторожность — верная профессиональная реакция, и в данном случае вполне применимо известное изречение: чрезвычайные заявления требуют чрезвычайных доказательств.

Почему же Перельман отверг Филдсовскую премию и отказался от награды Института Клэя? Это известно лишь ему самому, но вообще-то его никогда не интересовало признание такого рода, и он не раз говорил об этом. Он и раньше отказывался от премий и призов, правда, не таких престижных и крупных. Перельман с самого начала дал понять, что не хочет преждевременной известности. По иронии судьбы, именно это стало одной из причин, по которым специалисты не спешили высказывать свое мнение. Но, по правде говоря, не было ни единого шанса на то, что средства массовой информации *не заметят* его работу. Много лет математическое сообщество всеми силами старалось заинтересовать этой темой газеты, радио и телевидение, и нет смысла жаловаться на то, что эти усилия увенчались успехом, или ждать, что СМИ пропустят самую громкую математическую сенсацию со времен Великой теоремы Ферма. Но Перельман смотрел на все это иначе и спрятался в раковину. Поступило предложение — и оно остается в силе — использовать премиальные деньги на образовательные или иные цели, если он согласится. До сих пор ответа на это предложение нет.

11

Не могут они все быть легкими

Задача P/NP

В настоящее время математики используют компьютеры для решения самых разных задач, даже великих, и не считают это чем-то из ряда вон выходящим. Компьютеры хороши в числовых расчетах, но математика — это далеко не только «суммирование», так что ввести задачу в компьютер, как правило, очень непросто. Часто самое сложное — это преобразовать ее в такой вид, в каком ее можно решить путем компьютерных расчетов, и даже в этом случае компьютер иногда сопротивляется. Поэтому и в наше время решения многих великих задач находятся без или почти без участия компьютеров. Примерами тому — Великая теорема Ферма и гипотеза Пуанкаре.

В тех случаях, когда компьютеры использовались при решении великих задач (к примеру, теоремы о четырех красках или гипотезы Кеплера), они эффективно выступали в роли прислуги или помощников. Но иногда роли меняются, и математика становится служанкой компьютерной науки. Большая часть работы по первоначальному проектированию компьютеров шла в математическом ключе. Значительную роль в ней сыграла связь между булевой алгеброй — алгебраическим выражением формальной логики — и коммутационными схемами, разработанными, в частности, инженером Клодом Шенноном, создателем теории информации. Сегодня компьютеры и в практическом, и в теоретическом аспекте опираются на широкое использование многих самых разных областей математики.

Одна из задач тысячелетия по версии Института Клэя лежит в пограничной области между математикой и информатикой. В данном случае ситуацию можно рассматривать двояко: то ли информатика находится на службе у математики, то ли наоборот. На самом же деле требуется, да и развивается нечто другое, более сбалансированное, — партнерство. Задача касается компьютерных алгоритмов — математических скелетов, из которых вырастают компьютерные программы. Принципиальное значение здесь имеет концепция эффективности алгоритма: за сколько вычислительных шагов будет получен результат при определенном количестве входных данных. Практически эффективность говорит о том, сколько времени потребуется компьютеру на решение задачи заданного размера.

Слово «алгоритм» восходит к Средним векам, когда Мухаммад ибн-Муса аль-Хорезми написал один из первых трудов по алгебре. Еще раньше Диофант ввел в обращение элементы, которые мы сегодня прочно связываем с алгеброй: символы. У него они, однако, использовались как сокращения, а методы решения уравнений были представлены при помощи конкретных — хотя и типичных — примеров. Там, где мы сегодня написали бы что-нибудь вроде «$x + a = y$, следовательно, $x = y - a$», Диофант написал бы: «Положим $x + 3 = 10$, тогда $x = 10 - 3 = 7$» — и считал бы, что читатели должны сами сообразить, что эта идея будет работать и для любых других чисел вместо 3 и 10. Он готов был объяснить свой иллюстративный пример с применением символов, но никогда не стал бы оперировать символами как таковыми. Аль-Хорезми подробно разработал эту общую рекомендацию. Он сделал это при помощи слов, а не символов, но общая идея была верной, и именно он считается отцом алгебры. Более того, сам этот термин образован из названия его книги: она называлась «Краткая книга о вычислении посредством восполнения и противопоставления» («Аль-китаб аль-мухтасар фи хисаб аль-джебр ва-ль-мукабала»). Слово «аль-джебр», от которого и пошла алгебра, при этом означало «восполнение» и имело отношение к решению уравнений. А слово «алгоритм» произошло от средневековой латинизированной версии прозвища аль-Хорезми (т. е. «из Хорезма») — Алгорисмус — и означает в настоящее время специализированный математический процесс решения задачи, который при достаточно длительном ожидании гарантирует нахождение ответа на нее.

Традиционно в математике задача считалась решенной, если для нее в принципе можно было записать алгоритм, ведущий к ответу. Слово «алгоритм» использовалось редко: математики предпочитали представлять, скажем, формулу решения — частный случай алгоритма на языке символов. При этом было не слишком важно, может ли эта формула быть применена на практике: она сама по себе являлась решением. Но появление компьютеров изменило эту ситуацию. Формула, слишком сложная для ручных вычислений, вполне могла оказаться применимой, если привлечь на помощь компьютер. Зато ситуации, когда формула оказывалась слишком сложной даже для компьютера, стали вызывать раздражение, а такое тоже иногда случалось: конечно, любой алгоритм можно было попытаться просчитать на компьютере, но иногда расчет шел слишком медленно и не позволял получить ответ. Поэтому внимание ученых сместилось к поиску эффективных алгоритмов. И математики, и компьютерщики были кровно заинтересованы в получении алгоритмов, которые действительно позволяли бы за разумный промежуток времени получить ответ.

Если алгоритм имеется, относительно несложно оценить, сколько времени (измеряемого числом необходимых вычислительных операций) потребуется на решение задачи при определенном количестве входных данных. Это может требовать усилий и технических навыков, но зато вам известно, о каком именно процессе идет речь и, по крайней мере в общих чертах, что он делает. Гораздо сложнее разработать эффективный алгоритм, если тот, с которого вы начали, неэффективен. Еще сложнее решить, насколько плохим или хорошим может быть наилучший с точки зрения эффективности алгоритм для данной задачи, — ведь для этого нужно рассмотреть все возможные варианты, а вам неизвестно, что они собой представляют.

Первые работы по этому вопросу привели к грубому, но удобному разделению алгоритмов на эффективные (в простом, но неточном смысле) и неэффективные. Если продолжительность расчетов с увеличением количества входных данных растет относительно медленно, данный алгоритм эффективен, а задача проста. Если же продолжительность расчетов с увеличением количества входных данных растет очень быстро, то данный алгоритм неэффективен, а задача сложна. Опыт подсказывает, что, хотя встречаются задачи, которые в этом смысле можно назвать простыми, большинство задач к таковым не относятся и являются сложными. В самом деле, если бы все математические задачи были простыми, математики остались бы без работы. Соответствующая задача тысячелетия заключается в том, чтобы строго доказать, что существует по крайней мере одна сложная задача или что, вопреки нашему опыту, все задачи являются простыми. Эта задача известна как задача P/NP, и никто пока не представляет, как ее нужно решать.

В главе 2 мы уже сталкивались с приближенной оценкой эффективности. Алгоритм относится к классу P, если он имеет полиномиальное время работы. Иными словами, если число шагов, которые необходимо сделать для получения ответа, пропорционально количеству входных данных в какой-либо постоянной степени (скажем, в квадрате или кубе). Такие алгоритмы эффективны в самом широком смысле. Если входные данные представлены числом, то их количество — это не само число, а количество знаков в нем. Причина в том, что количество информации, необходимой для представления числа, соответствует месту, которое оно занимает в памяти компьютера, а это место пропорционально количеству цифр. Задача относится к классу P, если существует алгоритм класса P, который ее решает.

Любой другой алгоритм (или задача) принадлежит к классу не-P, и большинство таких алгоритмов неэффективны. Среди них есть алгоритмы, время работы которых экспоненциально по отношению к входным данным, т. е. примерно равно некоему фиксированному числу в степени, соответствующей размеру входных данных. Такие алгоритмы относятся к классу E и определенно неэффективны.

Некоторые алгоритмы настолько эффективны, что выполняют работу намного быстрее, чем за полиномиальное время. К примеру, чтобы определить четность или нечетность числа, достаточно посмотреть на его последнюю цифру. Если (в десятичной записи) это цифра 0, 2, 4, 6 или 8, число — четное, в противном случае — нечетное. Весь алгоритм включает в себя не более шести шагов:

```
Последняя цифра — 0? Если да, то СТОП. Число четное.
Последняя цифра — 2? Если да, то СТОП. Число четное.
Последняя цифра — 4? Если да, то СТОП. Число четное.
Последняя цифра — 6? Если да, то СТОП. Число четное.
Последняя цифра — 8? Если да, то СТОП. Число четное.
СТОП. Число нечетное.
```

Итак, время выполнения программы ограничено шестью шагами, вне зависимости от размера входных данных. Такой алгоритм относится к классу алгоритмов «постоянного времени».

Расстановка слов в списке в алфавитном порядке представляет собой задачу класса P. Простейший способ выполнить эту задачу — это так называемый «пузырьковый» метод, получивший название потому, что слова, находящиеся ниже по списку, чем следует, при этом «всплывают» вверх, как пузырьки в стакане газировки. Алгоритм раз за разом просматривает список, сравнивает соседние слова и меняет их местами, если порядок не соответствует алфавитному. Пусть, к примеру, список вначале выглядит так:

РОГ ДОМ БОТ АКТ

Сначала происходит следующее:

ДОМ РОГ БОТ АКТ
ДОМ **БОТ РОГ** АКТ
ДОМ БОТ **АКТ РОГ**

Полужирным шрифтом выделены те слова, сравнение которых проводилось только что и которые были (или не были) переставлены. При втором проходе получаем:

БОТ ДОМ АКТ РОГ
БОТ **АКТ ДОМ** РОГ
БОТ АКТ **ДОМ РОГ**

Третий проход:

АКТ БОТ ДОМ РОГ
АКТ **БОТ ДОМ** РОГ
АКТ БОТ **ДОМ РОГ**

На четвертом проходе ничего не происходит, так что мы понимаем, что программа завершила работу. Обратите внимание, как слово АКТ постепенно всплывает вверх (т. е. к началу списка).

Если в списке четыре слова, алгоритм на каждом шагу проводит три сравнения, а всего шагов получается четыре. Если слов в списке n, то на каждом проходе проводится $n-1$ сравнение, а проходов необходимо n, так что всего требуется $n\ (n-1)$ шагов. Это чуть меньше, чем n^2, так что время работы программы полиномиально, более того, квадратично. Алгоритм может прекратить работу раньше, но в самом худшем случае, если окажется, что слова в списке стоят точно в обратном порядке, ему потребуется $n\ (n\ -1)$ шагов. Пузырьковый алгоритм сортировки очевиден и относится к классу P, но это далеко не самый эффективный алгоритм. Самый быстрый алгоритм сортировки — алгоритм сравнения — организован более хитро и выполняется за $n\log n$ шагов.

Простой алгоритм с экспоненциальным временем работы — алгоритм класса E — это, к примеру, задание «распечатать список всех n-значных двоичных чисел». В таком списке 2^n чисел, и на печать (и вычисление) каждого уходит примерно n шагов, так что полное время работы составляет приблизительно $2^n n$, что больше, чем 2^n, но меньше, чем 3^n для достаточно больших n. Однако это довольно глупый пример, поскольку медленным его делает не сложность вычислений, а всего лишь размер выходных данных, и позже это наблюдение окажется весьма важным.

Более типичный алгоритм класса E решает задачу о коммивояжере. Этот странствующий продавец должен посетить некоторое количество городов. Делать это он может в произвольном порядке. Какой путь следует избрать, чтобы суммарное расстояние оказалось минимальным? Наивный способ решения этой задачи состоит в том, чтобы выписать все возможные маршруты, рассчитать для каждого суммарное расстояние и найти минимальное из всех. Для n городов у нас получится

$$n! = n\ (n-1) \times (n-2) \times \ldots \times 3 \times 2 \times 1$$

маршрутов (читается «n факториал»). Эта величина растет быстрее, чем любая экспоненциальная величина[37]. Более эффективный метод, известный как динамическое программирование, позволяет решить задачу о коммивояжере за экспоненциальное время. Первый подобный метод — алгоритм Хелда–Карпа — находит кратчайший маршрут за $2^n n^2$ шагов; при достаточно больших n это опять же попадает в интервал между 2^n и 3^n.

Несмотря на то что эти алгоритмы «неэффективны», при помощи специальных уловок можно ускорить расчет в случае, если число городов велико по человеческим меркам, но не слишком велико для применения подобных хитростей. В 2006 г. Д. Эпплгейт, Р. Биксби, В. Шваталь и У. Кук решили задачу о коммивояжере для 85 900 городов. На середину 2012 г. это достижение все еще оставалось рекордным.

Приведенные примеры алгоритмов не просто иллюстрируют концепцию эффективности. Кроме этого, они помогают донести до читателя мысль о сложности нахождения улучшенных алгоритмов и еще большей сложности нахождения алгоритмов максимальной эффективности. Все известные алгоритмы для задачи о коммивояжере относятся к классу E экспоненциального времени, но это не означает, что эффективного алгоритма для нее не существует в принципе. Это лишь показывает, что мы пока такого алгоритма не нашли. И здесь возможны два варианта: мы не нашли лучшего алгоритма потому, что нам не хватает ума, или мы не нашли лучшего алгоритма потому, что его попросту не существует.

Можно вспомнить все ту же вторую главу. Пока команда Агравала не придумала свой алгоритм класса P для проверки на простоту, наилучший известный алгоритм принадлежал классу не-P. Тем не менее он тоже был достаточно хорош, считал за время $n\log n$ для n-значных чисел, а это, вообще говоря, лучше, чем показатели алгоритма Агравала–Каяла–Саксены, пока мы не достигаем чисел с 10^{1000} знаками. До открытия этого алгоритма мнения о статусе испытания на простоту разделялись. Некоторые специалисты считали, что это задача класса P и подходящий алгоритм рано или поздно будет найден. Другие были уверены, что этого не произойдет. Новый алгоритм возник практически ниоткуда: его породила одна из бесчисленных идей, которые можно было попробовать, и данная конкретная идея сработала. Это отрезвляющий прецедент: мы не знаем истинного положения вещей, не можем предсказать его заранее, и догадки лучших экспертов могут быть как верными, так и ошибочными.

Великая задача, которая нас в данный момент интересует, заключается в поиске ответа на более фундаментальный вопрос. Существуют ли сложные задачи? Могут ли все задачи оказаться простыми, если, конечно, приложить достаточно ума и сообразительности? На самом деле здесь есть одна тонкость, потому что мы уже видели одну несомненно сложную задачу: распечатку списка всех n-значных двоичных чисел. Я уже упоминал о том, что это глупый пример: сложность заключается не в расчетах, а в простой монотонной работе по распечатке очень длинного ответа. Нам известно, что никакие уловки здесь не помогут, поскольку ответ будет таким длинным *по определению*. Если бы он был короче, он не был бы ответом.

Чтобы поставить вопрос разумным образом, необходимо исключить подобные тривиальные примеры. Для этого введем еще один класс алгоритмов, класс NP. Это не класс не-P — это класс алгоритмов, работа которых занимает недетерминированное полиномиальное время. В переводе с математического это означает, что, сколько бы времени ни требовалось алгоритму на поиск ответа, *убедиться в верности этого ответа* мы можем за полиномиальное время. Задача поиска ответа может быть сложной, но если ответ найден, то существует простой способ проверки его корректности.

Слово «недетерминированный» здесь используется потому, что существует возможность решить NP-задачу при помощи просто вдохновенной догадки. Сделав это, можно проверить и убедиться, что ответ действительно верен (или нет). К примеру, если задача заключается в разложении на простые множители числа 11 111 111 111, то вы можете предположить, что одним из множителей является простое число 21 649. Пока это всего лишь догадка, однако ее легко проверить: достаточно разделить исходное число на 21 649 и посмотреть, что получится. Частное равняется 513 239 точно, без остатка. Таким образом, ваша догадка оказалась верной. А если бы я догадался, что делителем должно быть 21 647 — тоже простое число, то деление привело бы к ответу 513 286 с остатком 9069. Таким образом, догадка оказалась бы неверной.

В данном случае правильное предположение можно сделать только чудом или при помощи обмана (я, кстати, прежде чем высказывать «предположение», разложил 11 111 111 111 на простые множители). Но, по существу, мы хотим именно этого. Если бы наша догадка не была чудесной, то можно было бы превратить алгоритм класса NP в алгоритм класса P очень простым способом: нужно было бы делать предположения одно за другим до тех пор, пока одно из них не оказалось бы верным. Мой пример позволяет увидеть, что так не получится: понадобилось бы слишком много попыток. В самом деле, то, что мы пытаемся делать, это всего лишь «пробное деление» на все возможные простые числа до тех пор, пока одно из них не сработает. Из главы 2 мы знаем, что это далеко не лучший способ искать делители.

Класс NP исключает глупые примеры вроде уже упоминавшегося очень длинного списка. Если кто-то в порыве вдохновения выдаст список всех n-значных двоичных чисел, то экспоненциальное время уйдет не только на то, чтобы их распечатать, но и на то, чтобы их прочесть, и еще больше времени — на то, чтобы проверить список. Это потребовало бы громадных корректорских усилий. Класс P определенно входит составной частью в класс NP. Если ответ можно найти за полиномиальное время, да еще с гарантией его корректности, то это будет означать, что вы его уже проверили. Так что проверка автоматически может быть произведена за полиномиальное время. Если бы кто-то представил вам предполагаемый ответ, то вы могли бы просто прогнать весь алгоритм еще раз — это и стало бы проверкой.

Теперь мы можем сформулировать задачу тысячелетия. Превосходит ли класс NP по размеру класс P или они суть одно и то же? Или короче: равен ли класс P классу NP?

Если ответ «да», то это значит, что существует принципиальная возможность отыскать быстрые и эффективные алгоритмы для автоматического составления расписаний авиарейсов, оптимизации работы завода, выполнения миллиона других важных практических задач. Если ответ «нет», то у нас будет железная гарантия того, что все вроде бы сложные задачи на самом деле сложны, и мы сможем остановиться и не тратить больше времени на поиск быстрых алгоритмов для них. В том и другом случае мы выигрываем. А вот не знать, как в реальности обстоят дела, очень неприятно.

Математикам было бы гораздо легче жить, если бы ответ был «да», поэтому пессимист, живущий в каждом человеческом существе, не может не заподозрить, что на самом деле все не так просто и ответ, скорее всего, окажется «нет». В противном случае мы все получаем бесплатный бонус, который ничем не заработали и которого не заслуживаем. Я, правда, подозреваю, что большинство математиков предпочло бы, чтобы ответ оказался «нет», потому что в этом случае им была бы гарантирована работа до конца времен. Математики самоутверждаются, решая сложные задачи. В общем, по разным причинам большинство математиков и компьютерщиков ожидают, что ответ на вопрос «Совпадает ли P с NP?» будет «Нет». И мало кто ждет, что ответом на самом деле окажется «да».

Помимо этого, возможны еще два варианта. Не исключено, что можно доказать эквивалентность P и NP, не находя в реальности полиномиальных алгоритмов для каждой конкретной NP-задачи. Математике свойственно предлагать нам неконструктивные доказательства существования: они утверждают, что нечто существует, но не говорят, что оно собой представляет и как его найти. В качестве примеров можно назвать методы проверки на простоту, которые бодро сообщают нам, что данное число не является простым, но не называют ни одного конкретного делителя, или теоремы теории чисел, уверяющие нас, что решения некоего диофантова уравнения ограничены, т. е. не превосходят некоторого предела, но не называющие никакого конкретного ограничения. В конце концов, полиномиальный алгоритм может быть настолько сложным, что записать его, в принципе, невозможно. Тогда естественный пессимизм в отношении бесплатного сыра окажется оправдан даже при положительном ответе на вопрос.

Некоторые исследователи высказываются еще более резко: они считают, что вопрос может оказаться нерешаемым в рамках современной математики, ограниченной формальной логикой. Если так, то невозможно доказать ни да ни нет. Не потому, что мы слишком глупы, чтобы найти доказательство, а потому, что такового не существует. Эта идея появилась на свет в 1931 г., когда Курт Гедель выпустил кошку противоречивости охотиться в стаю философских голубей, населявших подвалы математики (он доказал, что некоторые заявления в арифметике неразрешимы). В 1936 г. Алан Тьюринг нашел неразрешимую задачу попроще — задачу об остановке машины Тьюринга. Всегда ли при заданном алгоритме существует доказательство либо того, что машина остановится, либо того, что она будет считать вечно? Как ни удивительно, ответ Тьюринга был «нет». Для некоторых алгоритмов не существует доказательства ни того ни другого. Не исключено, что задача P/NP окажется такой же. Это объяснило бы, почему никто не может ни доказать, ни опровергнуть соответствующее утверждение. Но никто не может также доказать или опровергнуть утверждение о том, что задача P/NP неразрешима. Может быть, ее неразрешимость сама по себе неразрешима…

Самый очевидный подход к задаче P/NP состоит в том, чтобы выбрать какую-нибудь задачу, о которой известно, что она относится к классу NP, предположить существование полиномиального алгоритма ее решения — и каким-то образом прийти к противоречию. Некоторое время математики пытались применить эту методику к различным задачам, но в 1971 г. Стивен Кук понял, что выбор задачи часто не играет никакой роли. С определенной точки зрения все подобные задачи — с точностью до некоторых технических особенностей — совершенно равноправны. Кук ввел понятие NP-полной задачи. Такая NP-задача обладает следующим свойством: если для ее решения существует алгоритм класса P, то *любая* NP-задача может быть решена при помощи алгоритма класса P.

Кук нашел несколько NP-полных задач, включая SAT — задачу о выполнимости булевых формул. В ней спрашивается, можно ли сделать заданное логическое выражение истинным при помощи подходящего выбора значений (истинности или ложности) его переменных. Кроме того, он получил более глубокий результат: задача SAT с дополнительными ограничениями (3-SAT) также является NP-полной. Здесь логическая формула должна быть записана в виде «A, или B, или C, или… или Z», где A, B, C…Z — логические формулы, содержащие по три переменные. Спешу добавить, что переменные не обязаны каждый раз быть одними и теми же. Большинство доказательств того, что та или иная задача является NP-полной, восходят к теореме Кука о 3-SAT.

Определение Кука подразумевает, что все NP-полные задачи существуют на равных основаниях. Доказать, что одна из них на самом деле относится к классу P, означает доказать, что к классу P относятся все такие задачи. Это открывает некоторые тактические возможности: может оказаться, что с некоторыми NP-полными задачами работать проще, чем с остальными. Но стратегически это означает, что с тем же успехом можно выбрать любую конкретную NP-полную задачу и работать именно с ней. Все NP-полные задачи ведут себя одинаково, и поэтому на любой из них можно моделировать все остальные. А любую NP-задачу можно конвертировать в частный случай NP-полной задачи при помощи процедуры «шифрования» — с использованием шифра, на применение которого требуется полиномиальное время.

Чтобы представить себе характер этой процедуры, рассмотрим типичную NP-полную задачу: поиск гамильтонова цикла в сети. Требуется найти замкнутый маршрут по ребрам сети, которые прошел бы через каждый узел (т. е. через каждую точку) ровно один раз. «Замкнутый» означает, что в конце концов маршрут возвращается в начальную точку. Размер входных данных здесь — это число ребер, меньшее или равное квадрату числа точек, поскольку каждое ребро соединяет две точки. (Считаем, что любую заданную пару соединяет не больше одного ребра.) Нам не известно ни одного алгоритма класса P, который решал бы эту задачу, но предположим — гипотетически, — что такой алгоритм существует. Теперь выберем какую-нибудь другую задачу и назовем ее задачей X. Пусть задача X может быть переформулирована в терминах поиска такого маршрута в некоей сети, связанной с задачей X. Если метод перевода данных задачи X в данные об этой сети и наоборот, может быть применен за полиномиальное время, то мы автоматически получаем алгоритм класса P для задачи X. Примерно так:

1. Переводим задачу X в задачу поиска гамильтонова цикла в связанной с задачей сети. Это можно сделать за полиномиальное время.
2. Находим такой цикл за полиномиальное время при помощи того самого гипотетического алгоритма для задачи с сетью.
3. Переводим полученный гамильтонов цикл обратно в решение задачи X, что опять же можно проделать за полиномиальное время.

Поскольку все три полиномиальных шага вместе можно проделать тоже за полиномиальное время, этот алгоритм относится к классу P.

Чтобы показать, как это работает, я рассмотрю менее амбициозную версию задачи о поиске гамильтонова цикла, где искомый маршрут не обязан быть замкнутым. В таком виде она называется задачей о поиске гамильтонова пути. Сеть может иметь гамильтонов путь и при этом не иметь цикла (см. пример на рис. 42 слева). Так что решение задачи о поиске гамильтонова цикла может не означать решения задачи о поиске гамильтонова пути. Однако задачу о поиске гамильтонова пути можно переформулировать в задачу о поиске гамильтонова цикла на близкой, но несколько иной сети. Для этого в сеть добавляется одна дополнительная точка, соединенная со всеми точками первоначальной сети, как на рис. 42 справа. Любой гамильтонов цикл в новой сети может быть превращен в гамильтонов путь: для этого достаточно исключить из него добавленный узел и два подходящих к нему ребра цикла. И наоборот, любой гамильтонов путь в первоначальной сети дает цикл в новой: достаточно просто соединить два конца пути с новой точкой. Это «превращение» задачи о поиске пути в задачу о поиске цикла вводит в сеть всего одну новую точку и по одному ребру на каждую точку первоначальной сети, так что эта процедура — и обратная ей тоже — выполняется за полиномиальное время.

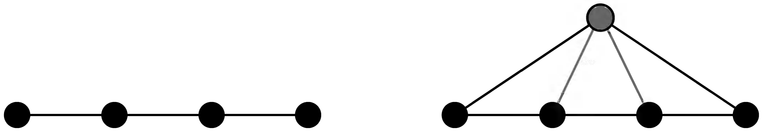


Рис. 42. Сеть с гамильтоновым путем (обозначен сплошной линией), но без гамильтонова цикла *(слева)*. Вводим дополнительную точку (обозначена серым) и четыре дополнительные линии, чтобы превратить гамильтонов путь в гамильтонов цикл (черная сплошная линия). Два серых ребра не входят в цикл, но нужны для строительства более крупной сети *(справа)*

Конечно, я здесь всего лишь зашифровал одну конкретную задачу, превратив ее в задачу о поиске гамильтонова цикла. Чтобы доказать, что такая задача является NP-полной, нам нужно проделать то же самое с любой NP-задачей. Это реально: первое доказательство нашел Ричард Карп в 1972 г. в знаменитой статье, где доказывалась NP-полнота 21 различной задачи.

Задача о коммивояжере является «почти» NP-полной, но здесь есть одна техническая сложность: мы не знаем, относится ли она к классу NP. Известно более 300 конкретных NP-полных задач в различных областях математики, включая логику, теорию графов, комбинаторику и оптимизацию. Доказать, что любая из них может (или не может) быть решена за полиномиальное время, означало бы доказать то же для всех них без исключения. Несмотря на богатство выбора, задача P/NP по-прежнему остается открытой. И я бы не удивился, если бы узнал, что она останется таковой и 100 лет спустя.

12

Потоковое мышление

Уравнение Навье–Стокса

Пять из семи задач тысячелетия, включая и три задачи, о которых мы уже говорили, относятся к чистой математике, хотя задача P/NP фундаментальна и для теории вычислительных систем. Оставшиеся две принадлежат к прикладной математике и современной математической физике. Задача из прикладной математики возникает из стандартного уравнения для потока жидкости — уравнения Навье–Стокса, названного в честь французского инженера и физика Клода-Луи Навье и ирландского математика и физика Джорджа Стокса. Уравнение Навье–Стокса — это уравнение в частных производных; следовательно, в нем учитывается скорость изменения характера потока как в пространстве, так и во времени. Большинство важнейших уравнений классической прикладной математики — это уравнения в частных производных (нам уже встречалось одно из таких уравнений — уравнение Лапласа); остальные — обыкновенные дифференциальные уравнения, учитывающие скорость изменения параметров только во времени.

В главе 8 мы видели, что движение тел в Солнечной системе определяется законом всемирного тяготения и законами динамики Ньютона. Эти законы связывают ускорение Солнца, Луны и планет с действующими на них гравитационными силами. Ускорение — это быстрота изменения скорости во времени, а скорость — характеристика изменения положения тела во времени. Это обычное дифференциальное уравнение. Как мы видели, решение таких уравнений может быть очень сложным делом. Как правило, решать дифференциальные уравнения в частных производных намного сложнее.

Если говорить о практических целях, то уравнения движения в Солнечной системе могут быть решены численно при помощи компьютеров. Это тоже непросто, но сегодня уже существуют хорошие методы. То же самое можно сказать и о решении в практических целях уравнений Навье–Стокса. Используемые при этом методики известны как вычислительная гидрогазодинамика и применяются для решения многих важных задач: конструирования самолетов, расчета аэродинамики автомобилей и даже в медицине (например, для расчета тока крови в организме человека).

Задача тысячелетия не просит математиков найти явные решения уравнения Навье–Стокса, поскольку это, по существу, невозможно. Не имеет она отношения и к численным методам решения этих уравнений, несмотря на всю их важность. Вместо этого в задаче требуется найти доказательство фундаментального теоретического свойства: *существования* решений. При заданном состоянии жидкости в определенный момент времени — при известных характеристиках ее движения — существует ли решение уравнения Навье–Стокса, верное для всего будущего времени начиная с рассматриваемого момента? Интуиция подсказывает, что ответ на этот вопрос должен быть «да», потому что данное уравнение — очень точная модель физики реальной жидкости. Однако с точки зрения математики вопрос существования решения не так очевиден, и это фундаментальное свойство для уравнения Навье–Стокса пока не доказано. А возможно, ответ все же «нет», и решения не существует.

Уравнение Навье–Стокса описывает, как меняется со временем в заданных условиях распределение скоростей в жидкости. О нем часто говорят во множественном числе как об уравнениях Навье–Стокса, но дела это не меняет. Множественное число отражает классический подход: в трехмерном пространстве скорость складывается из трех компонент; в классической теории на каждую компоненту приходится по одному уравнению, а всего их получается три. С современной точки зрения существует всего одно уравнение для *вектора* скорости (величины, которую характеризует не только размер, но и направление), но это уравнение приложимо к каждой из трех компонент скорости. На сайте Института Клэя используется классическая терминология, но здесь я буду следовать современной практике. Я говорю об этом заранее, чтобы избежать возможной путаницы.

Уравнение датируется 1822 г., когда Навье впервые записал уравнение в частных производных для потока вязкой — липкой — жидкости. Стокс внес свой вклад в 1842 и 1843 гг. Эйлер записал уравнение в частных производных для жидкости с нулевой вязкостью — совершенно не липкой — в 1757 г. Это уравнение тоже полезно, но большинство реальных жидкостей, включая воду и воздух, являются вязкими, поэтому Навье и Стокс модифицировали уравнение Эйлера таким образом, чтобы учесть это свойство. Они вывели примерно одинаковые уравнения независимо друг от друга, поэтому оно названо в честь их обоих. Навье сделал в процессе вывода несколько математических ошибок, но получил верный ответ, а у Стокса с математикой все было в порядке, и именно поэтому мы знаем, что ответ Навье верен, несмотря на ошибку. В самой общей форме уравнение применимо к сжимаемым жидкостям, таким как воздух. Однако существует и важный частный случай, при котором жидкость считается несжимаемой. Эта модель применима к таким жидкостям, как вода, которая под очень большим давлением все же сжимается, но лишь чуть-чуть.

Существует два способа составить математическое описание потока жидкости: можно либо описать маршрут движения каждой частицы жидкости со временем, либо описать скорость потока в каждой точке пространства и в каждый момент времени. Эти два описания связаны между собой: имея одно, можно (не без труда) вывести и второе. И Эйлер, и Навье, и Стокс использовали второй подход, потому что уравнение в этом случае получается гораздо более удобным и решаемым. Так что в их уравнениях фигурирует поле скоростей жидкости. В каждый конкретный момент времени поле скоростей точно определяет скорость и направление каждой частицы жидкости. По ходу времени это описание может меняться, именно поэтому в уравнении присутствуют скорости изменения параметров как в пространстве, так и во времени.

Уравнение Навье–Стокса имеет отличную физическую родословную. Оно основано на законах Ньютона, примененных к каждой крохотной частице (небольшой области) жидкости, и выражает в данном контексте закон сохранения импульса. Каждая частица движется, потому что на нее действуют силы, а закон движения Ньютона гласит, что ускорение частицы пропорционально действующей на нее силе. Основными силами являются трение, вызванное вязкостью, и давление. Присутствуют также силы, порожденные ускорением частицы. В соответствии с классической традицией уравнение описывает жидкость как бесконечно делимую массу. В частности, оно игнорирует дискретность атомной структуры жидкости в микромасштабе.

Уравнения сами по себе не имеют особой ценности: их надо еще научиться решать. Для уравнения Навье–Стокса решение означает расчет поля скоростей: скорости и направлении движения жидкости в каждой точке пространства в каждый момент времени. Уравнение налагает ограничения на эти величины, но не определяет их непосредственно. Вместо этого мы должны при помощи этого уравнения соотносить будущие скорости с текущими. Уравнения в частных производных, такие как уравнение Навье–Стокса, имеют много разных решений; точнее говоря, бесконечно много. И это неудивительно: жидкости способны течь очень по-разному: ток жидкости по капоту автомобиля отличается от тока жидкости по крылу самолета в полете. Существует два способа выбрать конкретный поток из бесконечного множества возможностей: используя либо начальные, либо граничные условия.

Начальные условия определяют поле скорости в какой-то конкретный момент времени; обычно его считают нулевым. Физически идея состоит в том, что если вам известно поле скорости в этот момент, то уравнение Навье–Стокса однозначно определяет это поле через очень короткий промежуток времени. Если для начала вы дадите жидкости толчок, она будет двигаться до тех пор, пока это не будет противоречить законам физики. Граничные условия более полезны в большинстве приложений, потому что начальные условия трудно обеспечить в реальной жидкости, да и вообще, они не слишком подходят для применения, скажем, в автомобильном дизайне. Там главное — форма машины. Вязкие жидкости прилипают к поверхностям. Математически это моделируется определением скорости на этих поверхностях, образующих границу занятой жидкостью области, а именно в ней уравнение действительно. К примеру, мы могли бы потребовать, чтобы скорость на границе была нулевой или наложить какое-то другое условие, которое наилучшим образом моделирует реальность.

Но даже в тех случаях, когда определены начальные или граничные условия, мы очень редко можем написать в явном виде формулу для поля скорости, потому что уравнение Навье–Стокса нелинейно. Сумма двух его решений, как правило, не является решением. Это, кстати, одна из причин, по которым задача трех тел из главы 8 настолько сложна, хотя это не единственная причина, ведь задача двух тел тоже нелинейна, но тем не менее решается в явном виде.

Для практических целей мы всегда можем решить уравнение Навье–Стокса на компьютере, представив поле скорости в виде набора чисел. Этот набор чисел можно очень наглядно представить в графическом виде и использовать для расчета величин, которые в первую очередь интересуют инженеров: к примеру, напряжений, возникающих в крыле самолета.

Поскольку компьютеры не умеют работать с бесконечным количеством чисел и не могут проводить вычисления с бесконечной точностью, нам приходится заменять реальный поток его дискретной аппроксимацией, т. е. набором чисел, представляющих поток в конечном числе точек и моментов времени. При этом очень важно, чтобы аппроксимация была достаточно качественной.

Обычный подход состоит в том, чтобы разделить пространство на большое число маленьких областей, образовав таким образом расчетную сетку. Скорость при этом вычисляется только в узлах этой сетки. Сама сетка может состоять из обычных квадратов (или кубов, если речь идет о трех измерениях), как шахматная доска, но для расчета автомобилей или самолетов она должна быть более сложной и иметь вблизи границы ячейки помельче, позволяющие уловить более тонкие детали происходящего. Сетка может быть динамической и менять форму с ходом времени. Обычно считается, что время идет дискретно, небольшими шагами, иногда одинаковыми, а иногда меняющими длительность в соответствии с ходом расчетов.

Основой большинства численных методов служит то, как «скорость изменения» определяется в дифференциальном исчислении. Предположим, некий объект сдвигается из одной точки в другую за очень короткий промежуток времени. Тогда скорость изменения положения — собственно скорость — есть изменение положения, деленное на время, которое на это потребовалось. При этом возникает небольшая ошибка, которая постепенно исчезает, по мере того как укорачивается промежуток времени. Поэтому мы можем аппроксимировать скорость изменения, входящую в уравнение Навье–Стокса, этим отношением изменения пространственного положения к изменению времени. В результате уравнение говорит нам, как провести известное начальное состояние — заданный набор скоростей — на один временной шаг вперед, в будущее. Примерно так же можно аппроксимировать решения, когда ситуация определяется граничными условиями. Существует также много хитрых способов добиться того же результата с большей точностью.

Чем мельче ячейки расчетной сетки и короче промежутки времени, тем точнее становится аппроксимация. Однако и вычислительный процесс при этом занимает больше времени. Так что приходится искать компромисс между точностью и скоростью. В широком смысле можно сказать, что приближенный ответ, полученный при помощи компьютера, скорее всего, окажется приемлемым, если в потоке нет значимых черт меньших по размеру, чем размер ячейки расчетной сетки. Существует два типа потока — ламинарный и турбулентный. В ламинарном потоке движение идет плавно, а слои жидкости скользят один относительно другого свободно, без трения. Здесь сетки с некрупными ячейками должно быть достаточно. Турбулентный поток — бурный и пенистый, и токи жидкости в нем перемешиваются чрезвычайно сложным образом. В подобных обстоятельствах дискретная сетка, какой бы частой она ни была, может не решить проблему.

Одна из особенностей турбулентного потока — присутствие вихрей, похожих на крохотные водовороты. Стандартное изображение турбулентности представляет собой каскад все более мелких вихрей. Большая часть деталей в таком потоке меньше по размеру, чем ячейка любой реальной сетки. Чтобы обойти эту сложность, инженеры в практических вопросах, касающихся турбулентных потоков, часто прибегают к статистическим моделям. Еще одна проблема заключается в том, что физическая модель жидкости может оказаться непригодной к рассмотрению турбулентных потоков, потому что вихри могут уменьшаться до атомных размеров. Однако сравнение численных расчетов и экспериментальных данных показывает, что уравнение Навье–Стокса — это очень реалистичная и точная модель. Она настолько хороша, что сегодня во многих инженерных приложениях целиком полагаются на вычислительную гидрогазодинамику: по сравнению с дорогими натурными экспериментами в аэродинамических трубах с использованием масштабных моделей компьютерные расчеты очень дешевы. Однако в вопросах, от которых зависит безопасность людей (к примеру, при проектировании самолетов), подобные экспериментальные проверки проводятся до сих пор.

Уравнения Навье–Стокса настолько точны, что приложимы, судя по всему, даже там, где с точки зрения физики вполне могли бы отказывать: в турбулентном потоке. По крайней мере так дело обстоит в случае, если уравнение может быть решено с достаточной точностью. Главная проблема здесь имеет практический характер: когда поток становится турбулентным, численные методы решения уравнения начинают поглощать громадное количество компьютерного времени. Кроме того, мельчайшие структуры всегда ускользают от «внимания» компьютера.

Математики очень не любят, когда информация о задаче, которой они располагают, основывается на каком бы то ни было приближении. Задача тысячелетия, связанная с уравнением Навье–Стокса, призвана решить один из ключевых теоретических вопросов. Если бы удалось найти ответ на него, интуитивное представление о том, что численные методы здесь прекрасно работают, получило бы мощнейшее подкрепление. Существует тонкая разница между приближениями, которые использует компьютер (он вносит в уравнение крохотные изменения), и точностью ответа (здесь речь идет о крохотных изменениях в решении). Можно ли сказать, что точный ответ на приближенно поставленный вопрос — то же самое, что приближенный ответ на точно поставленный вопрос? Иногда ответ бывает «нет». Точные данные о потоке жидкости с очень низкой вязкостью, к примеру, часто не совпадают с приближенными данными о потоке жидкости с нулевой вязкостью.

Есть один шаг к осмыслению подобных проблем, который настолько очевиден и прост, что его легко можно проглядеть: речь идет о доказательстве того факта, что точное решение существует. Ведь согласитесь, должно существовать *нечто,* аппроксимацией чего являются компьютерные расчеты.

Именно этим объясняется включение уравнения Навье–Стокса в число задач тысячелетия. Его официальное описание на сайте Института Клэя состоит из четырех отдельных задач. Решения любой из них будет достаточно для получения приза. Во всех четырех задачах жидкость считается несжимаемой. Вот эти задачи:

1. *Существование и гладкость решений в трех измерениях.* Здесь предполагается, что жидкость занимает бесконечное пространство целиком. При заданном первоначальном гладком поле скорости требуется доказать, что для любого положительного момента времени существует гладкое решение уравнения, совпадающее с заданным первоначальным полем.
2. *Существование и гладкость решений на трехмерном плоском торе.* Тот же вопрос, но теперь в предположении, что пространство представляет собой плоский тор — прямоугольник с отождествленными противоположными сторонами. В этой версии обходятся потенциальные проблемы, связанные с бесконечным пространством, о котором шла речь в первой версии; это пространство не соответствует реальности и может вызвать неправильное поведение системы по причинам, не имеющим непосредственного отношения к задаче.
3. *Опровержение существования решений в трех измерениях.* Опровергнуть пункт 1. Иными словами, найти начальное поле, для которого не существует гладкого решения в любой положительный момент времени, и доказать это утверждение.
4. *Опровержение существования решений на трехмерном плоском торе.* Доказать, что пункт 2 неверен.

Те же вопросы остаются открытыми и для уравнения Эйлера, которое, в сущности, эквивалентно уравнению Навье–Стокса, но не учитывает вязкости. За ответы на вопросы по уравнению Эйлера, однако, никто никаких призов не обещает.

Серьезная сложность здесь заключается в том, что рассматриваемый поток трехмерен. Существует аналогичное уравнение для жидкости, текущей по плоскости. Физически это может быть либо тонкий слой жидкости между двумя пластинами (считается, что они не вызывают трения), либо такой характер потока в трех измерениях, при котором жидкость движется совершенно идентичным образом вдоль системы параллельных плоскостей. В 1969 г. русский математик Ольга Ладыженская доказала, что для двумерного уравнения Навье–Стокса и двумерного уравнения Эйлера пункты 1 и 2 верны, а 3 и 4 — ложны.

Может показаться удивительным, но для уравнения Эйлера доказательство сложнее, чем для уравнения Навье–Стокса, хотя само уравнение проще, поскольку в нем не учитывается вязкость. Причина, надо сказать, весьма поучительная. Вязкость «снимает» вероятность того, что решение может привести к возникновению сингулярности, а та, в свою очередь, не позволит решению существовать в каждый момент времени. Если условие вязкости отсутствует, этого не происходит, что сказывается на математических характеристиках доказательства существования.

Ладыженская внесла еще одно важное дополнение в наши представления об уравнении Навье–Стокса, доказав не только, что решения существуют, но и что определенные вычислительные гидрогазодинамические модели аппроксимируют их с любой нужной нам точностью.

Задачи на приз тысячелетия относятся к несжимаемому потоку, поскольку хорошо известно, что сжимаемые потоки ведут себя отвратительно. В уравнениях движения самолета, к примеру, возникает множество проблем, если самолет движется в потоке воздуха быстрее звука. Это знаменитый «звуковой барьер», очень беспокоивший в свое время инженеров, которые работали над проектами сверхзвуковых истребителей. Эта проблема связана с хорошей сжимаемостью воздуха.

Если тело движется сквозь несжимаемую жидкость, оно расталкивает частицы этой жидкости в стороны со своего пути, как если бы это были шарики. Если частицы накапливаются, они замедляют тело. Но в сжимаемой жидкости, где существует предел скорости движения волн (а именно скорость звука), этого не происходит. На сверхзвуковых скоростях, вместо того чтобы расходиться в стороны, воздух скапливается перед самолетом, и его плотность там растет беспредельно. Результат — ударная волна. Математически это нарушение непрерывности давления воздуха, которое резко меняет значение на фронте ударной волны. Физически это звуковой удар: громкий хлопок. Ударная волна, если ее не учитывают, может повредить самолет, так что конструкторы волновались не зря. Однако скорость звука — не непреодолимый барьер, а всего лишь препятствие. Ее существование говорит о том, что уравнение Навье–Стокса для сжимаемой жидкости не обязательно имеет гладкие решения на всем диапазоне времен даже в двух измерениях. Так что в этом случае ответ известен заранее, и он отрицателен.

Математика ударной волны — большой раздел среди уравнений частных производных, несмотря на разрывы в решениях. Хотя уравнение Навье–Стокса само по себе не является хорошей физической моделью для сжимаемых жидкостей, математическую модель можно модифицировать, добавив к уравнениям дополнительные условия, которые помогут учесть ударную волну и нарушение непрерывности в ней. Но в потоке несжимаемой жидкости ударные волны не возникают, так что можно по крайней мере предположить, что в этом случае решения должны существовать для каждого момента времени, каким бы сложным (но обязательно гладким) ни было начальное состояние потока.

Кое-какие положительные результаты для трехмерного уравнения Навье–Стокса уже имеются. Если в начальном состоянии поток характеризуется достаточно маленькими скоростями, т. е. течет вяло и очень медленно, то и первое, и второе утверждения верны. Эти утверждения верны даже при больших скоростях — на протяжении некоторого ненулевого промежутка времени. Неизвестно, существует ли решение, верное для любого момента в будущем, но есть некоторый промежуток времени, на котором решение существует точно. Может показаться, что эту логику рассуждений можно повторять без конца, продвигая решение вперед во времени на небольшие промежутки и используя всякий раз конечный результат как новое начальное состояние. Проблема с подобным подходом заключается в том, что временны́е интервалы при этом могут уменьшаться настолько стремительно, что бесконечное число шагов будет укладываться в конечное время. К примеру, если каждый последовательный шаг продвигает решение на половину времени, достигнутого на предыдущем шаге, то весь процесс закончится за время $1+\frac{1}{2}+\frac{1}{4}+\frac{1}{8}+\ldots$, что равняется 2. Если решение прекращает существовать — в настоящее время это чисто гипотетическое предположение, но рассматривать его мы все же можем, то говорят, что решение, о котором идет речь, *разрушается*. Время, за которое это происходит, называется временем разрушения решения.

Так что в четырех задачах, по существу, спрашивается о том, могут ли решения разрушаться. Если не могут, верны утверждения 1 и 2; если могут — утверждения 3 и 4. Возможно, решения могут разрушаться в бесконечном пространстве, а на конечном плоском торе — не могут. Кстати говоря, если ответ на вопрос 1 положителен, то положителен ответ и на вопрос 2, потому что поток любой структуры на плоском торе можно интерпретировать как пространственно периодический поток в целом бесконечном пространстве. Речь идет о том, чтобы наполнить пространство копиями прямоугольника, о котором идет речь, и в каждом воспроизвести поток в точности той же структуры. Правила склеивания для тора гарантируют, что поток, пересекая эти плоские стыки, остается гладким. Аналогично если верно утверждение 4, то верно и утверждение 3 по той же причине. Мы просто делаем начальное пространство пространственно периодическим. Но, насколько мы сейчас в состоянии сказать, ответ на вопрос 2 может оказаться положительным даже при отрицательном ответе на вопрос 1.

Нам известен, однако, один поразительный факт, касающийся разрушения решений. Если существует решение с конечным временем разрушения, то максимальная скорость жидкости во всех точках пространства должна стать произвольно большой. Это могло бы произойти, к примеру, если бы сформировалась струя и скорость ее росла столь стремительно, что уже через конечный промежуток времени она улетела бы в бесконечность.

Это не чисто гипотетические возражения. Примеры подобного сингулярного поведения наблюдаются в некоторых других уравнениях классической математической физики. Замечательный пример можно найти в небесной механике. В 1988 г. Ся Чжихун доказал, что существует такая начальная конфигурация пяти материальных точек, или точечных масс, в трехмерном пространстве, где действует закон тяготения Ньютона, в которой четыре тела через конечный промежуток времени исчезают в бесконечности — тоже своего рода разрушение, а пятое переживает еще более значительные колебания. Ранее Джозеф Гервер указал, что пять тел на плоскости могут все раствориться в бесконечности за конечное время, но не смог завершить доказательство такого сценария. В 1989 г. он доказал, что разбегание такого рода определенно возможно на плоскости, если число тел достаточно велико.

Замечательно, что такое поведение возможно, ведь в подобных системах действует закон сохранения энергии. Конечно, если все тела движутся произвольно быстро, то полная кинетическая энергия системы должна расти? Ответ в том, что одновременно падает потенциальная энергия, а полная гравитационная потенциальная энергия материальной точки бесконечна. Должен сохраняться еще и момент импульса, но и это возможно, если некоторые из тел движутся все быстрее и быстрее по кругу все уменьшающегося диаметра.

Речь здесь идет о таком физическом явлении, как знаменитый эффект пращи, или гравитационный маневр, часто используемый при отправке исследовательских станций к далеким планетам Солнечной системы. Хороший пример — американский зонд «Галилео», в задачу которого входило долететь до Юпитера и исследовать эту гигантскую планету и ее многочисленные спутники. Зонд был запущен в 1989 г. и достиг цели в 1995 г. Путешествие длилось так долго, в частности, потому, что маршрут его был, мягко говоря, непрямым. Несмотря на то что орбита Юпитера находится дальше от Солнца, чем орбита Земли, «Галилео» в начале своего полета направился внутрь, к Венере. Он прошел вблизи Венеры, вернулся, чтобы пролететь мимо Земли, и отправился дальше в космос «взглянуть» на астероид Гаспра. Затем он вновь сблизился с Землей, *еще раз* обогнул нашу планету и наконец двинулся к Юпитеру. По пути он сблизился еще с одним астероидом, Идой, и обнаружил у него собственную крошечную луну — новый астероид, получивший название Дактиль.

Почему была выбрана такая извилистая траектория? От каждой встречи с планетой «Галилео» получал энергию и, следовательно, увеличивал скорость. Представьте себе, что космический зонд направляется к планете — не курсом столкновения, но так, чтобы пройти достаточно близко к поверхности и быстро развернуться за ней. После этого его должно вновь выбросить в дальний космос. Когда зонд проходит за планетой, они притягиваются друг к другу. Более того, они все время притягивались друг к другу, но на этой стадии полета сила притяжения становится максимальной и потому производит максимальное действие. Тяготение планеты как бы подталкивает зонд и придает ему дополнительную скорость. Суммарная энергия должна сохраняться, поэтому взамен зонд чуть замедляет движение планеты по орбите вокруг Солнца. Поскольку масса зонда очень мала, а масса планеты, напротив, очень велика, действием зонда на планету можно пренебречь. Действием планеты на зонд пренебречь нельзя: он может ускориться очень заметно.

«Галилео» прошел над поверхностью Венеры на высоте 16 000 км и получил прибавку скорости в 2,23 км/с. После этого он прошел в 960 км от Земли, а затем еще раз в 300 км, во второй раз добавив к своей скорости еще 3,7 км/с. Без этих маневров он не добрался бы до Юпитера, поскольку запускавшая его ракета не смогла бы направить его непосредственно туда. Первоначальный план, кстати говоря, предусматривал именно это: зонд предполагалось запустить на шаттле с кислородно-водородным разгонным блоком Centaur-G. Но катастрофа «Челленджера», когда космический челнок взорвался вскоре после старта, заставила отказаться от этого плана. Использование блока Centaur-G было запрещено. Пришлось воспользоваться для запуска «Галилео» менее мощным твердотопливным блоком IUS. Миссия была весьма успешна, среди ее научных результатов — наблюдение столкновения кометы Шумейкера–Леви с Юпитером, которое произошло в 1994 г., когда зонд был еще *на пути* к газовому гиганту.

Сценарий Ся учитывает и эффект пращи. Четыре планеты равной массы образуют две тесные пары, которые обращаются вокруг общих центров масс в двух параллельных плоскостях. Эти «ракетки», состоящие каждая из двух тел, играют в звездный теннис пятым, более легким телом, которое носится туда-сюда между ними по траектории, перпендикулярной плоскостям. Система устроена так, что всякий раз, когда этот «теннисный мячик» проходит мимо пары планет, эффект пращи ускоряет его и одновременно отталкивает пару планет прочь вдоль линии, соединяющей обе пары. Таким образом, «теннисный корт» с каждым ударом немного удлиняется, а игроки расходятся дальше. Энергия и импульс сохраняются в равновесии, поскольку две планеты, нанося «удар», придвигаются чуть ближе друг к другу и чуть ускоряют движение вокруг центра масс. При правильных начальных условиях пары планет расходятся все быстрее, и скорость их расхождения растет так стремительно, что они улетают в бесконечность за конечное время. При этом и «теннисный мяч» колеблется между ними все быстрее и быстрее. В сценариях разбегания Гервера тоже используется эффект пращи.

Но приложим ли этот фокус с исчезновением к реальным небесным телам? Нет, если подходить к вопросу буквально. В этих сценариях важно, чтобы тела были материальными точками. Для многих задач из небесной механики это достаточно разумное приближение, но не тогда, когда тела должны проходить на произвольно малых расстояниях друг от друга. Если бы тела конечных размеров действительно так делали, то рано или поздно они непременно столкнулись бы. Кроме того, релятивистские эффекты не позволили бы телам двигаться быстрее света и изменили бы закон гравитации. Во всяком случае начальные условия и дополнительное условие равенства некоторых масс в реальности, вероятно, никогда бы не выполнились. Тем не менее эти любопытные примеры показывают, что, хотя уравнения небесной механики, как правило, очень хорошо моделируют реальность, они могут иметь сложные сингулярности, которые не позволят решениям существовать в каждый момент времени. Не так давно ученые поняли, что в системе тройной звезды, где звезды движутся по сложным траекториям, эффект пращи может в какой-то момент выбросить одну из звезд наружу с большой скоростью. Так что вполне может оказаться, что галактику (а может, и межгалактическое пространство) бороздит несметное количество звезд-сирот — холодных, одиноких, нежеланных и невидимых, изгнанных братьями из своих систем.

Когда дифференциальное уравнение ведет себя так странно, что его решения через конечный промежуток времени лишаются всякого смысла, мы говорим, что возникает сингулярность. Описанная выше работа по задаче множества тел на самом деле посвящена различным типам сингулярности. В задаче тысячелетия, связанной с уравнением Навье–Стокса, спрашивается, могут ли сингулярности возникать в задачах с начальными условиями для жидкости, занимающей либо все пространство, либо плоский тор. Если сингулярность может сформироваться за конечное время, результатом, скорее всего, станет разрушение решения, разве что сингулярность разрешится позже сама собой, что представляется маловероятным.

Существует два основных подхода к этим вопросам. Можно попытаться доказать, что сингулярностей не возникает, а можно попытаться найти одну из них, подобрав подходящие начальные условия. В том и другом могут помочь численные решения: они могут предложить полезные общие свойства потоков, а могут дать кое-какие указания на возможную природу потенциальных сингулярностей. Однако в численных решениях потенциально теряется точность, поэтому к любым указаниям такого рода следует относиться с осторожностью и обосновывать их более строго.

В попытках доказать регулярность, т. е. отсутствие сингулярностей, ученые пытаются получить контроль над потоком при помощи целого ряда методов. Среди них сложные оценки величины тех или иных ключевых переменных или еще более абстрактные методики. Популярный подход предлагает воспользоваться так называемыми слабыми решениями, которые являются не потоками в точном смысле этого слова, а более общими математическими структурами с некоторыми свойствами потоков. Известно, к примеру, что набор сингулярностей любого слабого решения трехмерных уравнений Навье–Стокса всегда мал.

Уже исследованы многие сценарии, которые могли бы вести к сингулярностям. Так, в 1941 г. Андрей Колмогоров разработал стандартную на сегодня модель турбулентности в виде каскада бесконечно уменьшающихся вихрей. Он предположил, что на очень мелких масштабах все формы турбулентности выглядят исключительно похоже. Пропорции вихрей заданного размера, к примеру, подчиняются универсальному закону. В настоящее время известно, что по мере уменьшения вихри меняют форму, становятся длиннее и тоньше и образуют элементарные струйки. Из закона сохранения момента импульса следует, что завихренность — степень закрученности вихрей — должна возрастать. Этот процесс называется растягиванием вихрей, и именно такое поведение, в принципе, может привести к сингулярности: к примеру, если мельчайшие вихри растянулись бы и стали бесконечно длинными за конечное время, а завихренность в некоторых точках стала бы бесконечной.

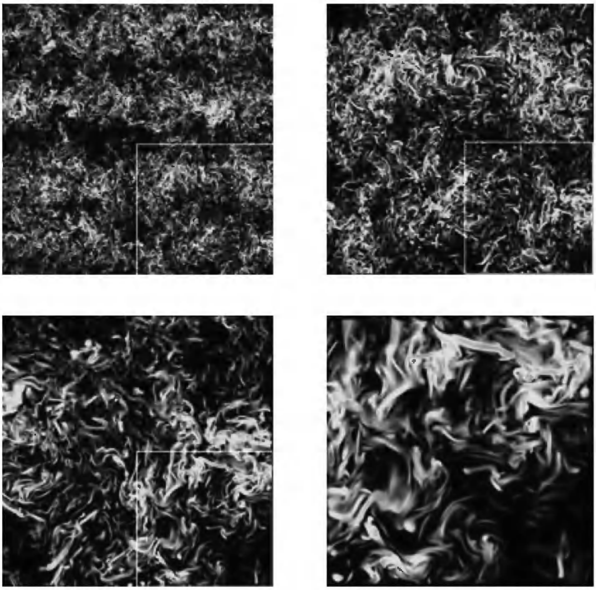


Рис. 43. Увеличенные изображения турбулентного потока, смоделированные компьютерной системой VAPOR

На рис. 43 можно увидеть сильно увеличенное изображение турбулентного потока, смоделированного Пабло Мининни и его коллегами с использованием программы VAPOR — платформы визуализации и анализа для океана, атмосферы и солнечных исследований. На изображениях видна интенсивность завихренности: насколько быстро вращается жидкость. Они иллюстрируют формирование вихревых струек (видны как длинные тонкие структуры) и показывают, что элементарные струйки могут собираться в пучки и образовывать более крупные структуры. Программа позволяет проводить моделирование на кубической решетке более чем с тремя миллиардами узловых точек.

В статье, посвященной этой теме и размещенной на сайте Института Клэя, Чарльз Фефферман написал:

> «Существует множество интереснейших задач и гипотез о поведении решений уравнений Эйлера и Навье–Стокса… Поскольку мы не знаем даже, существуют ли эти решения, наши представления о них находятся на очень примитивном уровне. Стандартные методы [из теории дифференциальных уравнений в частных производных] представляются недостаточными для решения этой задачи. Вместо этого нам, вероятно, требуются новые глубокие идеи».

Сложность потока на изображениях, подобных рис. 43, помогает представить себе трудности, с которыми, вероятно, столкнутся исследователи в поисках таких идей. Но математики не сдаются — они продолжают идти вперед, пытаясь отыскать простые принципы в видимой сложности.

13

Квантовая головоломка

Массовая щель

В нескольких километрах к северу от Женевы граница между Швейцарией и Францией делает резкий изгиб. На поверхности в этом месте видны лишь проселочные дороги и небольшие деревеньки. Но под землей, на глубине от 50 до 175 м, находится самый крупный на планете научный прибор. Это гигантский кольцевой туннель более 8 км в диаметре, соединенный с другим, меньшим (примерно вчетверо) туннелем. Большая его часть находится под территорией Франции, но две секции приходятся на Швейцарию. По туннелям проложено по паре труб, которые сходятся в четырех точках.

Это Большой адронный коллайдер стоимостью €7,5 млрд (около $9 млрд). С его помощью ученые ведут исследования на переднем крае физики элементарных частиц. Ключевой целью 10 000 ученых более чем из 100 стран, вместе работающих на этой экспериментальной установке, было найти бозон Хиггса или не найти его, если так устроена Вселенная. Частица эта нужна была, чтобы дополнить Стандартную модель элементарных частиц, согласно которой все во Вселенной состоит из элементарных частиц 17 разновидностей. В теории бозон Хиггса — то, что придает всем частицам массу.

В декабре 2011 г. два основных детектора Большого адронного коллайдера — ATLAS и CMS — независимо друг от друга обнаружили предварительные свидетельства существования бозона Хиггса с массой около 125 ГэВ (гигаэлектронвольт —

единица, которую в физике элементарных частиц используют для обозначения равно массы и энергии, поскольку то и другое эквивалентно). А 4 июля 2012 г. ЦЕРН — Европейский центр ядерных исследований, управляющий Большим адронным коллайдером, — объявил заинтересованной аудитории ученых и научных журналистов, что Вселенная высказалась в пользу Хиггса. Обе группы собрали большое количество дополнительных данных, и вероятность того, что данные показали случайную флуктуацию, а не присутствие новой частицы с хиггсовскими характеристиками, составляет теперь менее одной двухмиллионной. Именно такую степень уверенности традиционно требуют в физике элементарных частиц, прежде чем открывать шампанское.

Только дальнейшие эксперименты позволят с уверенностью сказать, обладает ли новая частица всеми теми свойствами, которые теоретически должны быть у бозона Хиггса. К примеру, теория предсказывает, что спин бозона Хиггса должен быть равен нулю, а на момент объявления наблюдаемые данные указывали на значение 0 либо 2. Может также оказаться, что «настоящий» бозон Хиггса состоит из других, меньших частиц или что это всего лишь первая ласточка из нового семейства хиггсоподобных частиц. В итоге либо существующая модель элементарных частиц будет подтверждена и закреплена, либо мы получим новую информацию, которая со временем позволит нам разработать другую, лучшую теорию.

Последняя из семи задач тысячелетия тесно связана со Стандартной моделью и бозоном Хиггса. Это центральный вопрос квантовой теории поля — математической области, в рамках которой изучается физика элементарных частиц. Его еще называют гипотезой «щели» в спектре масс, и он устанавливает конкретный нижний предел для возможной массы элементарной частицы. Это лишь одна репрезентативная задача, выбранная из целой серии крупных нерешенных вопросов в этой новейшей области математической физики, и она связана как с вопросами из самых передовых разделов чистой математики, так и с давней мечтой физиков — теорией, которая объединила бы две важнейшие физические теории: общую теорию относительности и квантовую теорию поля.

В классической ньютоновой механике фундаментальными физическими понятиями являются пространство, время и масса. Пространство считается трехмерным и евклидовым, время — одномерная величина, не зависящая от пространства, а масса указывает на присутствие вещества. Массы изменяют свое положение в пространстве под действием различных сил, и скорость, с которой меняется их положение, измеряется относительно времени. Ньютонов закон движения описывает, как ускорение тела (скорость изменения его скорости, которая, в свою очередь, отражает скорость изменения его позиции) соотносится с его массой и приложенной к нему силой.

Классические теории пространства, времени и вещества поднялись на максимальную высоту в уравнениях электромагнетизма, предложенных Джеймсом Максвеллом. Эта элегантная система уравнений объединила две силы природы, которые раньше рассматривались исключительно по отдельности. Оказалось, что вместо отдельных явлений электричества и магнетизма существует единое электромагнитное поле. Это поле пронизывает все пространство, как если бы Вселенная была наполнена какой-то невидимой жидкостью. В каждой точке пространства мы можем измерить величину и направление этого поля, как будто эта жидкость течет по математическим законам. Конечно, для некоторых целей электромагнитное поле можно разбить на два компонента, два поля: электрическое и магнитное. Но переменное магнитное поле порождает электрическое поле, и наоборот, так что, когда дело доходит до динамики, оба они должны рассматриваться совместно, как единое более сложное поле.

Эта удобная и уютная картина физического мира, в котором фундаментальные физические концепции описывали предметы и явления, воспринимаемые с помощью человеческих органов чувств, резко изменилась в самом начале XX в. Именно тогда физики начали понимать, что на очень маленьких масштабах, слишком мелких для тогдашних микроскопов и вообще каких бы то ни было средств наблюдения, вещество выглядит совсем не так, как считалось прежде. Физики и химики начали принимать всерьез странную теорию, возникшую более 2000 лет назад и восходящую к философствованиям древнего грека Демокрита и индийских ученых. Идея заключалась в том, что, хотя мир, судя по всему, сделан из бесчисленного множества различных материалов, все вещество на самом деле построено из крохотных частичек — атомов. Слово «атом» по-гречески означает «неделимый».

Химики XIX в. нашли немало косвенных свидетельств существования атомов: элементы, соединяясь вместе и образуя более сложные молекулы, делают это в очень конкретных соотношениях, часто близких к целым числам. Джон Дальтон сформулировал свои наблюдения в виде закона кратных пропорций и предложил в качестве объяснения атомы. Если каждое химическое соединение состоит из фиксированного числа атомов разных видов, то такое соотношение появится автоматически. К примеру, нам сегодня известно, что каждая молекула двуокиси углерода состоит из двух атомов кислорода и одного атома углерода, так что атомы в этом веществе всегда будут присутствовать в отношении два к одному. Однако есть и сложности: разные атомы имеют разную массу, а многие элементы существуют в виде молекул, состоящих из нескольких атомов — к примеру, молекула кислорода состоит из двух атомов кислорода. И если вы не понимаете, что происходит, то решите, что атом кислорода вдвое массивнее, чем на самом деле. Кроме того, некоторые распространенные элементы на самом деле представляют собой смесь разных «изотопов» — атомных структур. К примеру, хлор существует в природе в виде смеси двух стабильных форм, известных как хлор-35 и хлор-37 в соотношении примерно 76 и 24% соответственно. Так что наблюдаемый «атомный вес» хлора равен 35,45. Зарождающаяся атомная теория интерпретировала это как «атом хлора состоит из тридцати пяти с половиной атомов водорода». А это означало, что атом не является неделимым. Век XX уже начался, а большинство ученых по-прежнему считало, что принятие атомной теории — слишком решительный шаг, чтобы сделать его, основываясь на таких данных.

Некоторые ученые, в первую очередь Максвелл и Людвиг Больцман, продвинулись дальше других и были убеждены, что газы — это тонко распределенные наборы молекул и что молекулы получаются при соединении атомов. Большинство их коллег убедило, судя по всему, данное Эйнштейном объяснение броуновского движения — хаотичного движения крохотных частиц взвеси, видимых под микроскопом. Эйнштейн решил, что эти подергивания вызываются столкновениями с хаотично движущимися молекулами жидкости. Он также провел кое-какие численные расчеты, подтвердившие эту точку зрения. Жан Перрен в 1908 г. подтвердил эти предположения экспериментально. Возможность видеть действие предполагаемых неделимых частиц вещества и делать на основании увиденного численные предсказания оказалась для ученых гораздо более убедительной, чем философские рассуждения и занятная нумерология. В 1911 г. Амедео Авогадро[1] разобрался в проблеме изотопов, и существование атомов было признано окончательно.

Пока все это происходило, кое-кто из ученых начал понимать, что атомы вовсе не являются неделимыми. Они обладают структурой, и от них можно отбивать маленькие кусочки. В 1897 г. Джозеф Томсон, экспериментируя с катодными лучами, открыл, что атомы можно заставить испускать еще более мелкие частицы, электроны. И не только это: оказалось,

---

[1] Изотопы были открыты в 1906–1907 гг. при исследовании продуктов радиоактивного распада тяжелых элементов. Название «изотоп» было предложено в 1910 г. Фредериком Содди. Амедео Авогадро (1776–1856) ровно на 100 лет раньше, в 1811 г., предложил метод определения масс молекул исходя из пропорций, в которых вещества вступают в химические реакции. — *Прим. пер.*

что атомы разных элементов испускают одни и те же частицы. При помощи магнитного поля Томсон показал, что электроны несут отрицательный электрический заряд. Но атом электрически нейтрален, так что в нем должна быть какая-то часть, обладающая положительным зарядом. Обдумав это, Томсон предложил модель атома, известную как «пудинг с изюмом»: атом похож на положительно заряженный пудинг с отрицательно заряженными электронами-изюминками внутри. Но в 1909 г. Эрнест Резерфорд, один из бывших студентов Томсона, провел эксперимент и продемонстрировал, что большая часть массы атома сосредоточена возле его центра. Пудинги такими не бывают.

Как можно экспериментально прозондировать такую крохотную область пространства? Представьте себе участок земли, на котором могут быть здания и другие сооружения, а может и не быть ничего. Вам не позволяется входить на эту территорию, к тому же вокруг темно, хоть глаз выколи, и ничего не видно. Однако у вас есть винтовка и неограниченный запас патронов. Вы можете стрелять наугад в направлении участка и отслеживать направление, в котором пули из него вылетают. Если участок напоминает пудинг с изюмом, то большая часть пуль пролетит насквозь по прямой. Если вам придется время от времени уворачиваться от пуль, срикошетивших прямо на вас, то можно будет сделать вывод, что впереди находится что-то довольно твердое. Наблюдая за тем, как часто пули вылетают с участка под тем или иным углом, вы сможете оценить размеры твердого объекта.

Пулями Резерфорда стали альфа-частицы — ядра атомов гелия, а участком земли для него служила тончайшая золотая фольга. Работа Томсона показала, что электроны-изюминки обладают очень малой массой, так что почти вся масса атома должна была приходиться на сам пудинг. Если бы в пудинге не было уплотнений, то бо́льшая часть альфа-частиц должна была бы пролетать насквозь. Лишь некоторые частицы могли отклоняться от своего пути, и то ненамного. Вместо этого оказалось, что небольшая, но заметная часть альфа-частиц отклонялась на достаточно большие углы, что явно не соответствовало картине пудинга. Резерфорд предложил другую метафору, которой мы часто пользуемся и сегодня, несмотря на существование более современных моделей. Речь идет о планетарной модели атома. Атом подобен Солнечной системе, предположил Резерфорд: в нем есть громадное центральное ядро, «солнце» системы, а вокруг ядра, подобно планетам, обращаются электроны. Поэтому атом, как и Солнечная система, по большей части представляет собой пустое пространство.

Резерфорд пошел дальше и нашел доказательства того, что ядро состоит из двух различных типов частиц: протонов, несущих положительный заряд, и нейтронов с нулевым зарядом. Массы тех и других очень близки и примерно в 1800 раз превосходят массу электрона. Таким образом, атомы не только не являются неделимыми, но и состоят из еще более мелких субатомных частиц. Эта теория объясняет целочисленную нумерологию химических элементов: оказывается, подсчитывается не что-нибудь, а количество протонов и нейтронов. Кроме того, она объясняет изотопы: добавление или удаление нескольких нейтронов изменяет массу атома, но сохраняет его суммарный нулевой заряд и число электронов, равное числу протонов. Химические свойства атома определяются в основном его электронами. К примеру, хлор-35 содержит 17 протонов, 17 электронов и 18 нейтронов; хлор-37 — 17 протонов, 17 электронов и 20 нейтронов. Атомная масса 35,45 возникает потому, что природный хлор представляет собой неравную смесь этих двух изотопов.

В начале XX в. появилась и новая теория, применимая к веществу в масштабе субатомных частиц. Она получила название «квантовая механика», и после ее появления физика принципиально изменилась и уже никогда не будет прежней. Квантовая механика предсказала множество новых явлений, которые затем удалось пронаблюдать в лаборатории, и существование новых элементарных частиц. Она также помогла понять прежде не поддававшиеся объяснению явления. Наконец, она изменила наши представления о Вселенной, поскольку классический ее образ, несмотря на великолепную согласованность со всеми предыдущими наблюдениями, оказался неверен. Человеческие органы чувств плохо приспособлены для восприятия реальности на фундаментальном уровне.

В классической физике вещество состоит из частиц, а свет представляет собой волну. В квантовой механике свет тоже частица, фотон; и наоборот, вещество (к примеру, электроны) может иногда вести себя как волна. Прежнее четкое деление на волны и частицы не то чтобы размывается, а вовсе исчезает, сменяясь корпускулярно-волновым дуализмом. Если воспринимать все буквально, планетарная модель атома работала не слишком хорошо, поэтому вскоре появился новый образ. Электроны не обращаются вокруг ядра, как планеты вокруг Солнца, а образуют размытое облако с центром в ядре — облако вероятностей, а не чего-то конкретного. Плотность облака в некоторой точке соответствует вероятности обнаружить в данной точке электрон.

Итак, помимо протонов, нейтронов и электронов физики знали еще одну субатомную частицу — фотон. Вскоре появились и другие. Кажущееся нарушение закона сохранения энергии побудило Вольфганга Паули предложить коллегам исправить положение — постулировать существование нейтрино, невидимой и практически необнаружимой новой частицы, которая объяснила бы утечку энергии. Необнаружимость частицы, однако, оказалась неполной, что позволило в 1956 г. подтвердить ее существование. После этого как будто распахнулись шлюзы. Пионы, мюоны, каоны посыпались как из рога изобилия (последние были открыты в результате наблюдения космических лучей). Появилась новая дисциплина — физика элементарных частиц, и первым ее рабочим инструментом стал метод Резерфорда, позволявший проводить зондирование на тех невероятно малых масштабах, о которых шла речь: чтобы выяснить, как устроен тот или иной объект, нужно бомбардировать его разными «снарядами» и смотреть на результат. Началось строительство и использование все более масштабных ускорителей частиц — по существу, орудий, стреляющих теми самыми пробными снарядами. Стэнфордский линейный ускоритель имел длину 3 км. Чтобы не строить ускорителей длиной в целый континент, их стали изгибать и замыкать в круг, чтобы частицы могли беспрерывно двигаться по ним, одновременно набирая колоссальные скорости. Это серьезно усложнило технологию, поскольку частицы при движении по кругу излучают энергию, но с этим научились справляться.

Первым результатом этих трудов стал растущий каталог элементарных вроде бы частиц. Энрико Ферми так выразил свое разочарование: «Если бы я мог запомнить названия всех этих частиц, я был бы ботаником». Однако время от времени в квантовой теории появлялись новые идеи, и список вновь менялся: предлагались очередные мельчайшие частицы, чтобы объединить уже наблюдавшиеся структуры.

Вначале квантовая механика описывала отдельные волноподобные или частицеподобные явления, но никто не мог вразумительно описать квантово-механический аналог поля. Однако игнорировать этот пробел было невозможно, потому что частицы, описываемые квантовой механикой, могут взаимодействовать и взаимодействуют с полями, которые на тот момент квантовой механикой не описывались. Представьте, что кто-то захотел бы выяснить, как движутся планеты Солнечной системы, притом что ньютоновы законы движения (описывающие, как движутся массы под действием сил) были бы известны, а вот его же закон тяготения (объясняющий, что представляют собой эти силы) — нет.

Но помимо частиц была и другая причина стремиться прояснить вопрос с полями. Благодаря корпускулярно-волновому дуализму то и другое теснейшим образом связано. По существу, частица — это скомканный кусочек поля, а поле — это море плотно упакованных частиц. Эти две концепции неразделимы. К несчастью, разработанные к тому моменту методы были основаны на том, что частицы похожи на крохотные точки, и никак не распространялись на поля. Невозможно просто согнать множество частиц в одно место и назвать то, что получилось, полем, потому что частицы *взаимодействуют* друг с другом.

Представьте толпу людей… к примеру, в поле. Может быть, они собрались там послушать рок-концерт. Если посмотреть из пролетающего вертолета, толпа людей похожа на жидкость, хлюпающую в поле — часто буквально, как, к примеру, на фестивале в Гластонбери: известно, что поле там превращается в море грязи. Внизу, на земле, становится ясно, что на самом деле жидкость — это бурлящая масса отдельных частиц: людей. Или, возможно, тесных небольших групп людей, таких как несколько гуляющих вместе друзей, которые представляют собой неделимую единицу, или как группа незнакомых людей, объединенных общей целью — к примеру, походом в бар. Но невозможно точно смоделировать толпу, просто сложив воедино поведение отдельных людей (то, как они вели бы себя в одиночестве). Направляясь к бару, одна группа преграждает путь другой, группы сталкиваются и перемешиваются. Разработка эффективной квантовой теории поля напоминает моделирование поведения толпы, в которой роль людей выполняют локализованные квантовые волновые функции.

К концу 1920-х гг. физики убедились (в частности, при помощи подобных рассуждений), что, как бы трудна ни была задача, квантовую механику придется расширять, чтобы она могла описывать не только частицы, но и поля. Естественной отправной точкой для этого стало электромагнитное поле. Необходимо было каким-то образом квантовать и электрический, и магнитный его компоненты, т. е. переписать его характеристики на языке квантовой механики. Но тут возникали сложности. Математический аппарат квантовой механики был незнаком и к тому же выглядел крайне нефизически. То, что можно было увидеть и измерить, уже не выражалось добрыми старыми числами, а соответствовало операторам гильбертова пространства: математическим правилам, разработанным для работы с волнами. Эти операторы нарушали обычные постулаты классической механики. При перемножении двух чисел результат не зависит от их порядка; к примеру, 2 × 3 и 3 × 2 — это одно и то же. Это свойство сложения, известное как коммутативность, нарушается для многих пар операторов — примерно так же, как надеть сначала носки, а затем ботинки, не то же самое, что сначала надеть ботинки, а затем носки. Числа — существа пассивные, а вот операторы — активны. Действие, которое вы произведете первым, подготавливает сцену для дальнейших событий.

Коммутативность — очень приятное математическое свойство. Его отсутствие раздражает и мешает, поэтому, в частности, квантование поля оказалось такой хитрой задачей. Тем не менее она решаема. Электромагнитное поле удалось квантовать в несколько этапов. Начался этот процесс с теории электрона Дирака (1928 г.), а завершили его Синъитиро Томонага, Джулиан Швингер, Ричард Фейнман и Фримен Дайсон в конце 1940-х — начале 1950-х гг. Получившаяся в результате теория стала называться квантовой электродинамикой.

Точка зрения, использованная при разработке этой теории, давала подходы к методу, который мог бы применяться и более широко. В основе его лежала идея, восходившая непосредственно к Ньютону. Пытаясь решить уравнения, связанные с законом Ньютона, ученые открыли несколько полезных общих принципов, известных как законы сохранения. Дело в том, что при движении системы массивных тел некоторые величины остаются неизменными. Самая известная из них — энергия, которая бывает двух видов: кинетическая и потенциальная. Кинетическая энергия определяется тем, насколько быстро движется тело, а потенциальная — представляет собой работу, проделанную определенными силами. Когда камень падает со скалы, он как бы обменивает потенциальную энергию, связанную с тяготением, на кинетическую. Говоря обычным языком, он падает и ускоряется. Кроме этого, сохраняются такие величины, как импульс, равный произведению массы на скорость, и момент импульса, связанный со скоростью вращения тела. Сохраняющиеся величины связывают различные переменные, используемые для описания системы, и таким образом уменьшают их число. Это очень полезно при решении уравнений, как мы уже видели в главе 8, где речь шла о задаче двух тел.

К началу XX в. ученые разобрались в том, откуда взялись законы сохранения. Эмми Нетер доказала, что каждая сохраняющаяся величина соответствует непрерывной группе симметрий в уравнениях. Симметрия — это математическое преобразование, при котором уравнения не меняются. Все симметрии образуют группу с операцией «провести одно преобразование, затем другое». Непрерывная группа — это группа симметрий, определенная единственным действительным числом. К примеру, вращение вокруг заданной оси есть симметрия, и угол вращения может задаваться любым действительным числом, поэтому вращения — на все возможные углы — вокруг заданной оси образуют непрерывную группу. Из сохраняющихся величин с этой симметрией связан момент импульса, или вращательный момент. Точно так же сохранение импульса связано с непрерывной группой перемещений в заданном направлении. А как насчет энергии? Ее сохранение связанно с временны́ми симметриями — уравнения неизменны в любой момент времени.

Попытавшись унифицировать фундаментальные силы природы, физики убедились, что ключ к единой теории — именно симметрии. Первым такая унификация удалась Максвеллу, который соединил электричество и магнетизм в единое электромагнитное поле. Максвелл сделал это без привлечения симметрии, но вскоре стало ясно, что в его уравнениях присутствует особый вид симметрии, которого прежде никто не замечал: калибровочная симметрия. Создавалось впечатление, что она может стать стратегическим рычагом, при помощи которого ученым удастся открыть путь к более общим квантовым теориям поля.

Вращение и перенос — глобальные симметрии: они равно применимы в любой точке пространства и времени. Вращение вокруг определенной оси поворачивает на один и тот же угол каждую точку пространства. Не таковы калибровочные симметрии: это местные симметрии, они могут меняться от одной точки пространства к другой. В случае электромагнетизма местные симметрии — это смена фазы. Колебания электромагнитного поля в определенной точке обладают амплитудой (это размах колебаний) и фазой (это момент, в который колеблющаяся величина достигает своего максимума). Если взять решение уравнений поля Максвелла и в каждой точке поменять фазу, то получится другое решение (если, конечно, вы внесете в описание поля соответствующее компенсирующее изменение, включающее местный электромагнитный заряд).

Калибровочные симметрии ввел в обращение Герман Вейль в безуспешной попытке добиться дальнейшей унификации электромагнетизма и общей теории относительности, т. е. электромагнитных и гравитационных сил. Название появилось в результате недопонимания: он считал, что правильная местная симметрия должна означать изменение пространственного масштаба, т. е. «калибровку». Из этой идеи ничего не получилось, но логика квантовой механики заставила Владимира Фока и Фрица Лондона предложить другой тип местной симметрии. Квантовая механика формулируется с использованием не только действительных, но и комплексных чисел, и каждая квантовая волновая функция имеет комплексную фазу. Значимые местные симметрии вращают фазу на любой угол на комплексной плоскости. В принципе, эта группа симметрий включает в себя все вращения, но в комплексных координатах все они представляют собой «унитарные трансформации» (U) в пространстве с одним комплексным измерением (1), поэтому группа, сформированная этими симметриями, обозначается как U(1). Формальные обозначения здесь не просто математическая игра: они позволили физикам записать, а затем и решить уравнения для заряженных квантовых частиц, движущихся в электромагнитном поле. Именно благодаря этому Томонага, Швингер, Фейнман и Дайсон разработали первую релятивистскую квантовополевую теорию электромагнитных взаимодействий: квантовую электродинамику. Симметрия калибровочной группы U(1) играла в их работах фундаментальную роль.

Следующий шаг, объединивший квантовую электродинамику с теорией слабого ядерного взаимодействия, сделали в 1960-е гг. Абдус Салам, Шелдон Глэшоу, Стивен Вайнберг и другие ученые. К электромагнитному полю с его калибровочной симметрией U(1) они добавили поля, связанные с четырьмя элементарными частицами — так называемыми бозонами W^+, W^0, W^- и B^0. Калибровочные симметрии такого поля, по существу, вращают комбинации этих частиц, порождая другие их комбинации; эти симметрии образуют другую группу, получившую обозначение U(2) — унитарные (U) трансформации в двумерном комплексном пространстве (2), являющиеся также специальными (S) — простое формальное условие. Иными словами, полная калибровочная группа — это U(1) × SU(2), где знак × указывает на то, что две группы действуют независимо на двух разных полях. Результат, получивший название теории электрослабых взаимодействий, потребовал введения сложного математического новшества. Группа U(1) в квантовой электродинамике коммутативна: два проведенных последовательно симметричных преобразования дают один и тот же результат, в каком бы порядке они ни проводились. Это свойство сильно упрощает всю математику, но для группы SU(2) не работает. Так впервые была применена некоммутативная калибровочная теория.

Сильное ядерное взаимодействие вступает в игру при рассмотрении внутренней структуры таких частиц, как протоны и нейтроны. Толчком к большому прорыву в этой области послужила интересная математическая закономерность, наблюдаемая в одном конкретном классе частиц, известных как адроны. Эта закономерность, известная как «восьмеричный путь», вдохновила ученых на создание теории квантовой хромодинамики. Теория постулировала существование скрытых частиц, названных кварками, и использовала их в качестве базовых компонент для целого зоопарка адронов.

Согласно Стандартной модели, все во Вселенной состоит из 16 по-настоящему элементарных частиц, существование которых подтверждено экспериментами на ускорителях. Плюс 17-я частица, поисками которой в настоящее время занят Большой адронный коллайдер. Из частиц, известных еще Резерфорду, ранг элементарных сохранили только две: электрон и фотон. Протон и нейтрон, напротив, состоят из кварков. Это название пустил в оборот Марри Гелл-Ман, позаимствовав его из романа Джеймса Джойса «Поминки по Финнегану». Гелл-Ман хотел, чтобы слово quark произносилось как «корк», однако фраза из романа Джойса, в которой оно встречается: «Три кварка для мастера Марка!» — подразумевает, что слово quark должно рифмоваться с именем Марк. Тем не менее Гелл-Ман нашел способ обосновать свое намерение. Сегодня в английском языке распространены оба варианта произношения.

В Стандартной модели предполагается существование шести кварков, объединенных попарно. Названия кварков довольно забавны: верхний/нижний, очарованный/странный, истинный/прелестный. Кроме того, модель предусматривает шесть лептонов, тоже парных: электрон, мюон и таон (который чаще называют по старинке тау-частицей) и соответствующие им нейтрино. Все эти 12 частиц называют фермионами — в честь Энрико Ферми. Частицы удерживаются вместе силами четырех типов: это гравитация, электромагнетизм, сильное и слабое ядерные взаимодействия. Оставив в стороне гравитацию, которую до сих пор не удалось полностью согласовать с квантовой картиной мира, получаем три силы. В физике элементарных частиц действие сил осуществляется посредством обмена частицами, которые их «переносят» или «передают». Традиционная аналогия — теннисисты, которых удерживает в пределах площадки их обоюдное внимание к мячу. Фотон переносит электромагнитное взаимодействие, Z- и W-бозоны переносят слабое ядерное взаимодействие, а посредством глюона передается сильное ядерное взаимодействие. Технически глюон переносит «цветное» взаимодействие, или взаимодействие между кварками, удерживающее их вместе, и, соответственно, сильное взаимодействие, которое мы наблюдаем в результате. Протон состоит из двух верхних кварков и одного нижнего; нейтрон — из двух нижних и одного верхнего. В каждой из этих частиц именно глюоны удерживают кварки на месте. Четыре перечисленных переносчика взаимодействий обобщенно называют бозонами, в честь Шатьендраната Бозе. Разница между фермионами и бозонами очень существенна: у них разные статистические свойства. На рис. 44 слева можно увидеть итоговый каталог предположительно элементарных частиц. На рис. 44 справа показано, как собрать протон и нейтрон из кварков.

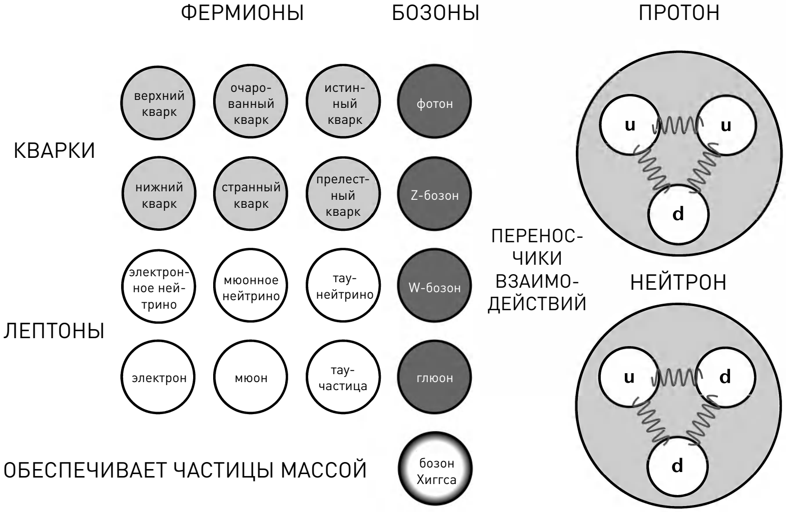


Рис. 44. 17 частиц Стандартной модели *(слева)*. Как собрать протон и нейтрон из кварков *(справа)*. Протон = два u-кварка (верхних кварка) + + один d-кварк (нижний кварк) *(справа вверху)*. Нейтрон = один u-кварк + + два d-кварка *(справа внизу)*

Бозон Хиггса завершает картину и объясняет, почему остальные 16 частиц Стандартной модели обладают ненулевыми массами. Он назван в честь Питера Хиггса — одного из тех физиков, которым принадлежала первоначальная идея. Кроме него, в работе над теорией, связанной с бозоном Хиггса, участвовали Филип Андерсон, Франсуа Энглер, Роберт Браут, Джеральд Гуральник, Карл Хаген и Томас Киббл. Бозон Хиггса — это воплощенное в частице гипотетическое квантовое поле — поле Хиггса — с необычным, но очень важным свойством: в вакууме оно не равно нулю. Это поле действует на остальные 16 частиц модели, заставляя их вести себя так, будто они обладают массой.

В 1993 г. Дэвид Миллер, отвечая на вызов британского министра науки Уильяма Уолдгрейва, предложил замечательную аналогию. Представьте себе многолюдную вечеринку. Гости равномерно распределены по залу, и тут входит почетный гость (отставной премьер-министр). Сразу же вокруг него собирается толпа народу. Гость движется по залу, кто-то из других гостей присоединяется к группе, кто-то отходит. Толпа сопровождающих придает почетному гостю дополнительную массу, ему теперь трудно остановиться. Это и есть механизм Хиггса. А теперь представьте, что по залу неожиданно разносится какой-то слух, и люди собираются послушать новости. Эта группа — бозон Хиггса. Миллер тогда добавил: «Может оказаться, что механизм Хиггса и поле Хиггса пронизывают всю Вселенную, а бозона Хиггса не существует. Следующее поколение коллайдеров прояснит этот вопрос». Судя по всему, вопрос с бозоном Хиггса действительно прояснился, а вот поле Хиггса требует дополнительных исследований.

Квантовая хромодинамика — это еще одна калибровочная теория, на этот раз с калибровочной группой SU(3). Как можно понять из обозначения, на этот раз преобразование действует на трехмерном комплексном пространстве. Из этого выводится унификация электромагнетизма, слабого и сильного взаимодействий. Предполагается, что существует три квантовых поля, по одному на каждое взаимодействие, с калибровочными группами U(1), SU(2) и SU(3) соответственно. Комбинация всех трех полей дает Стандартную модель с калибровочной группой U(1) × SU(2) × SU(3). Строго говоря, симметрии SU(2) и SU(3) приблизительны; считается, что они становятся точными при очень высоких энергиях. Поэтому их действие на частицы, составляющие ткань нашего мира, соответствует нарушенным симметриям — следам структуры, которые сохраняются в идеальной, полностью симметричной системе, подвергнувшейся небольшим возмущениям.

Все три группы содержат непрерывные семейства симметрий: одно семейство U(1), три — SU(2) и восемь — SU(3). Со всеми ними связаны различные сохраняющиеся величины. Симметрии ньютоновой механики, как обычно, обеспечивают сохранение энергии, импульса и момента импульса. Калибровочные симметрии U(1) × SU(2) × SU(3) свидетельствуют о сохранении различных «квантовых чисел», характеризующих частицы. Квантовые числа аналогичны таким величинам, как спин и заряд, но в отношении к кваркам. Здесь можно услышать такие названия, как цветовой заряд, изоспин или гиперзаряд. Наконец, в связи с U(1) сохраняются еще кое-какие величины: речь идет о квантовых числах шести лептонов, таких как электронное число, мюонное число и тау-число. В результате всего этого получается, что симметрии уравнений Стандартной модели объясняют через теорему Нетер все существенные физические переменные элементарных частиц.

Для нашей истории важны общая стратегия и результат. Чтобы унифицировать физические теории, нужно отыскать и унифицировать их симметрии. Затем нужно придумать подходящую теорию, в которой фигурировала бы объединенная группа симметрий. Я не говорю, что это простой и прямолинейный процесс, — технически это очень сложно. Но до сих пор квантовая теория поля развивалась именно так, и только одно из четырех фундаментальных физических взаимодействий — гравитация — пока выпадает из общей картины.

Теорема Нетер не только объясняет основные физические переменные, связанные с элементарными частицами, — именно так были открыты многие базовые симметрии. Исходя из квантовых чисел, которые удалось установить путем наблюдений или логических рассуждений, физики пытались выяснить, какими симметриями в этом случае должна обладать модель. Затем они составляли подходящие уравнения с этими симметриями и убеждались, что эти уравнения достаточно точно отражают реальность. В данный момент последний этап требует подбора величин 19 параметров — чисел, которые необходимо подставить в уравнения для получения количественных результатов. Девять из девятнадцати — это массы конкретных частиц: всех шести кварков, а также электрона, мюона и тау-частицы. Остальные параметры более технические: например, углы смешивания и фазовые связи. Семнадцать параметров известны из экспериментов, но два — все еще нет: они описывают до сих пор гипотетическое *поле* Хиггса. Однако сегодня есть все шансы измерить их, поскольку физики знают, где их искать.

Уравнения, которые используются в этих теориях, относятся к общему классу калибровочных теорий поля, известных как теория Янга–Миллса. В 1954 г. Янг Чжэньнин и Роберт Миллс попытались разработать калибровочные теории для объяснения сильного взаимодействия и связанных с ним частиц. Первые попытки закончились неудачей: после квантования поля выяснилось, что массы частиц при этом должны быть нулевыми. В 1960 г. Джеффри Голдстоун, Ёитиро Намбу и Джованни Йона-Лазинио нашли способ обойти эту проблему: они начали с теории, предсказывавшей безмассовые частицы, но затем модифицировали ее, постулировав нарушение некоторых симметрий. Иными словами, слегка изменили уравнения, введя в них новые асимметричные условия. Когда при помощи той же идеи модифицировали теорию Янга–Миллса, то получившиеся уравнения очень хорошо легли и в электромагнитную теорию, и в квантовую хромодинамику.

Янг и Миллс предположили, что калибровочная группа является специальной унитарной группой. К частицам применимы группы SU(2) и SU(3), специальные унитарные группы для двух или трех комплексных измерений, но вообще-то этот математический аппарат работает для любого числа измерений. Их теория в лоб атакует сложную, но неизбежную математическую проблему. В одном отношении электромагнитное поле отличается обманчивой простотой: его калибровочные симметрии коммутативны. В отличие от большинства квантовых операторов фазы можно менять в любом порядке. Но физики-то работали с квантовой теорией поля для субатомных частиц. Там калибровочная группа не коммутативна, что очень затрудняет квантование уравнений.

Добиться успеха Янгу и Миллсу помогло схематическое представление взаимодействий частиц, предложенное Ричардом Фейнманом. Любое квантовое состояние может быть представлено как суперпозиция бесчисленных взаимодействий частиц. К примеру, даже в вакууме есть пары частиц и античастиц, которые на мгновение возникают из небытия и тут же исчезают вновь. Простое столкновение двух частиц порождает умопомрачительный танец, в котором промежуточные частицы появляются и исчезают, мечутся взад и вперед, расщепляются и сливаются. Спасает лишь сочетание двух подходов. Уравнения поля для каждой конкретной фейнмановской диаграммы можно проквантовать, а затем сложить все отдельные вклады и представить себе полный эффект взаимодействия. Более того, самые сложные диаграммы встречаются редко и потому их вклад в общую сумму невелик. Тем не менее здесь есть серьезная проблема. Сумма, если рассматривать ее буквально, бесконечна. Янг и Миллс нашли способ перенормировать расчет таким образом, чтобы исключить бесконечное число слагаемых, которые, по идее, не должны много значить. Осталась конечная сумма, и ее величина очень точно соответствовала реальности. При первом знакомстве эта методика казалась почти непостижимой, но сегодня в ней многое прояснилось.

В 1970-е гг. к делу подключились математики. Майкл Атья обобщил теорию Янга–Миллса на большой класс калибровочных групп. Математика и физика начали подпитываться друг от друга. Работа Эдварда Уиттена и Натана Зайберга над топологическими квантовыми теориями поля породила концепцию суперсимметрии, в которой каждая известная частица имеет «суперсимметричного» партнера: электрону соответствует селектрон, кваркам — скварки. Это предположение упростило математику и позволило сделать кое-какие физические предсказания. Однако никому еще не удалось наблюдать хотя бы одну из этих новых частиц, а некоторые из них, вероятно, уже должны были появиться в экспериментах на Большом адронном коллайдере. В математической ценности этих идей никто не сомневается, а вот их непосредственное значение в физике пока под вопросом. Тем не менее они помогли многое прояснить в теории Янга–Миллса.

Квантовая теория поля — один из наиболее динамично развивающихся передовых рубежей математической физики, поэтому Институт Клэя захотел включить в группу задач тысячелетия что-нибудь из этой области. Выбрали проблему массовой щели. Речь в ней идет о важном математическом вопросе из физики элементарных частиц. Применение полей типа Янга–Миллса для описания элементарных частиц в терминах сильного ядерного взаимодействия сильно зависит от особого квантового свойства, известного как массовая щель. В теории относительности частица, летящая со скоростью света, приобретает бесконечную массу, если только ее масса покоя не равна нулю. Щель в спектре масс позволяет квантовым частицам иметь конечную ненулевую массу, несмотря на то что связанные с ними классические волны движутся со скоростью света. Если массовая щель существует, то любое состояние, не являющееся вакуумом, обладает энергией, превышающей энергию вакуума по крайней мере на некоторую фиксированную величину. Иными словами, существует ненулевой нижний предел массы частицы.

Эксперименты подтверждают существование массовой щели, и компьютерное моделирование уравнений тоже говорит в пользу этой гипотезы. Однако мы не можем считать, что модель соответствует реальности, а затем использовать данные экспериментов (т.е. реальность) для проверки математических свойств модели, потому что в этом случае логика зацикливается. Необходимо теоретическое доказательство. Ключевым шагом здесь стало бы строгое доказательство того, что квантовые версии теории Янга–Миллса существуют. В классическом (неквантовом) ее варианте ученые уже довольно хорошо разобрались, но квантовый аналог осложняется проблемой перенормировки — теми самыми бесконечностями, избавляться от которых приходится при помощи математических уловок.

Один многообещающий подход начинается с того, что непрерывное пространство превращают в дискретную пространственную решетку и записывают для решетки уравнение, аналогичное уравнению Янга–Миллса. Затем главное — показать, что по мере того, как решетка становится все мельче, постепенно приближаясь к сплошной среде, этот аналог сходится к четко определенному математическому объекту. На основании физической интуиции можно сделать вывод о некоторых необходимых его свойствах, и если бы эти свойства удалось установить строго, то можно было бы доказать и существование подходящей квантовой теории Янга–Миллса. Гипотеза о массовой щели требует более детальных представлений о том, как решетчатые теории аппроксимируют эту гипотетическую теорию Янга–Миллса. Так что существование этой теории и гипотеза массовой щели тесно взаимосвязаны.

На этом этапе все и застопорилось. В 2004 г. Майкл Дуглас составил отчет о состоянии проблемы, в котором написал: «Насколько мне известно, в последние годы в этом вопросе не было никаких прорывов. В частности, хотя в области теорий поля для низких размерностей достигнут некоторый прогресс, мне неизвестно, о каком бы то ни было существенном прогрессе в строительстве математически строгой квантовой теории Янга–Миллса». Судя по всему, это утверждение справедливо до сих пор.

В некоторых смежных задачах, однако, наблюдался более впечатляющий прогресс, и не исключено, что это поможет пролить свет и на интересующий нас вопрос. Частные случаи квантовой теории поля, известные как двумерные сигма-модели, разрешимы легче, и для одной такой модели гипотеза массовой щели уже доказана. Суперсимметричные квантовые теории поля, в которых фигурируют гипотетические суперпартнеры обычных элементарных частиц, отличаются некоторыми математическими свойствами, которые по существу, делают перенормировку ненужной. Физики, такие как Эдвард Уиттен, продвигаются к решению соответствующих задач в суперсимметричном случае. Можно надеяться, что некоторые из разработанных ими методик, возможно, подскажут новые пути решения первоначальной задачи. Но каковы бы ни были физические следствия и как бы ни разрешился в конце концов вопрос существования массовой щели, наработки, уже сделанные в этой области, безусловно, обогатили математику новыми важными понятиями и инструментами.

14

Диофантовы мечты

Гипотеза Берча–Свиннертон-Дайера

В главе 7 мы уже встречались с «Арифметикой» Диофанта, и я упоминал о том, что 6 из 13 ее книг дошли до нас в греческих копиях. Примерно в 400 г. н. э., когда древнегреческая цивилизация уже давно находилась в упадке, лидерство в математической науке захватили Аравия, Китай и Индия. Арабские ученые перевели классические греческие работы, и сегодня мы знаем многие из них лишь по этим переводам. Именно в арабском мире развивались идеи Диофанта. Четыре арабские рукописи, найденные в 1968 г., могут быть переводами неизвестных до сих пор книг «Арифметики».

В какой-то момент в конце X в. персидский математик аль-Караджи задал вопрос, который, вполне возможно, приходил в голову и самому Диофанту: какие целые числа могут возникать в качестве одинаковой разности между тремя рациональными квадратами, образующими арифметическую последовательность? К примеру, целые квадраты 1, 25 и 49 имеют общую разность 24. Иными словами, $1 + 24 = 25$ и $25 + 24 = 49$. Аль-Караджи жил примерно между 953 и 1029 гг., и он, в принципе, мог иметь доступ к списку «Арифметики», однако первый известный перевод сделал Абу-л-Вафа в 998 г. Леонард Диксон, автор краткой истории теории чисел в трех томах, предположил, что эта задача могла возникнуть незадолго до 972 г. в арабской рукописи неизвестного автора.

На алгебраическом языке задача звучит так: для каких целых d существует рациональное число x такое, что $x - d$, x и $x + d$ являются полными квадратами? Ее можно сформулировать и иначе, хотя эквивалентность формулировок неочевидна: какие целые числа могут представлять собой площадь прямоугольного треугольника с рациональными сторонами? Иными словами, если a, b и c рациональны и $a^2 + b^2 = c^2$, то какие целые значения возможны для величины $ab/2$? Целые числа, удовлетворяющие этим эквивалентным условиям, называют конгруэнтными. Термин не имеет отношения к остальным случаям использования слова «конгруэнтный» в математике, и современного читателя это может несколько сбивать с толку. Его происхождение объясняется ниже.

Некоторые числа не являются конгруэнтными: к примеру, можно доказать, что 1, 2, 3 и 4 неконгруэнтны. С другой стороны, 5, 6 и 7, напротив, конгруэнтны. В самом деле, площадь треугольника со сторонами 3, 4, 5 равна $3 \times 4/2 = 6$, что доказывает конгруэнтность числа 6. Чтобы доказать конгруэнтность числа 7, заметим, что треугольник со сторонами $(24/5)^2$, $(35/12)^2$ и $(337/60)^2$ также прямоугольный и его площадь равна 7. К числу 5 я вернусь чуть позже. Рассматривая числа поочередно, одно за другим, мы получим длинный список конгруэнтных чисел, но вряд ли прольем много света на их природу. Никакое количество конкретных примеров не докажет, что какое-то конкретное целое число *не является* конгруэнтным. Несколько сотен лет никто не мог сказать, конгруэнтно число 1 или нет.

Сегодня мы знаем, что эта задача далеко выходит за рамки всего, что Диофант хотя бы в принципе мог решить. Более того, этот обманчиво простой вопрос полностью не разрешен до сих пор. Максимум, что нам удалось получить, — характеризация конгруэнтных чисел, открытая Джеральдом Таннеллом в 1983 г. Идея Таннелла позволяет получить алгоритм для определения, может ли данное целое число возникать в соответствующих ситуациях при помощи расчета его представлений в виде двух различных комбинаций квадратов. При небольшой изобретательности этот расчет годится для достаточно больших целых чисел. Эта характеризация имеет всего один серьезный недостаток: никто еще не доказал, что она верна. Ее адекватность зависит от решения одной из задач тысячелетия — гипотезы Берча–Свиннертон-Дайера. Эта гипотеза предлагает критерий, при котором эллиптическая кривая имеет конечное число рациональных точек. Мы уже встречали эти диофантовы уравнения в главе 6 (гипотеза Морделла) и главе 7 (Великая теорема Ферма). В этой главе мы еще раз увидим, какую выдающуюся роль они играют в теории чисел.

Самая ранняя из европейских работ, посвященных этим вопросам, принадлежит перу Леонардо Пизанского. Нам Леонардо по прозвищу Фибоначчи известен прежде всего благодаря последовательности странных чисел, которую он, судя по всему, изобрел. Числа эти возникали в ходе решения арифметической задачи о размножении каких-то невероятных кроликов. Вот числа Фибоначчи:

0 1 1 2 3 5 8 13 21 34 55 89…

В этом ряду каждое число после двух первых представляет собой сумму двух предыдущих чисел. Отцом Леонардо был таможенный чиновник по имени Боначчо, и знаменитое прозвище означает «сын Боначчо». У нас нет никаких данных о том, что это прозвище использовалось при жизни Леонардо. Считается, что его придумал французский математик Гийом Либри в XIX в. Как бы то ни было, числа Фибоначчи широко известны и обладают множеством поразительных свойств. Они даже фигурируют в крипто-конспирологическом триллере Дэна Брауна «Код да Винчи».

Леонардо представил свои числа Фибоначчи в учебнике по арифметике «Книга счета» (Liber Abbaci), написанном в 1202 г. Основной целью учебника было привлечь внимание европейцев к придуманной арабами новой форме записи чисел, в основе которой лежали десять цифр от 0 до 9, и продемонстрировать ее универсальность. Сама идея десятичной записи уже достигла Европы через текст аль-Хорезми 825 г., названный в латинском переводе «Об индийском счете» (Algoritmi de Numero Indorum), но книга Леонардо стала первой из тех, что были написаны именно для того, чтобы способствовать внедрению десятичной системы в Европе. Значительная часть книги посвящена практической арифметике, в первую очередь операциям по обмену денег. Кроме этого, Леонардо написал еще одну книгу. Она не так известна, хотя во многих отношениях является непосредственным преемником диофантовой «Арифметики». Называется она «Книга квадратов» (Liber Quadratorum).

Подобно Диофанту, Леонардо представлял общие методики через конкретные примеры. Один из них основывался на вопросе аль-Караджи. В 1225 г. Пизу посетил император Фридрих II. Он был наслышан о Леонардо и его математических занятиях и, судя по всему, решил, что будет забавно объявить математический турнир и посмотреть на него в деле. В то время подобные публичные состязания были обычным делом. Участники задавали друг другу вопросы. В команду императора входили Джованни из Палермо и магистр Теодор. В команду Леонардо входил только сам Леонардо. Команда императора попросила Леонардо найти такой квадрат, который остался бы квадратом, если вычесть из него или прибавить к нему 5. Как обычно, все числа должны были быть рациональными. Иными словами, соперники хотели, чтобы Леонардо доказал, что 5 — число конгруэнтное, отыскав конкретное рациональное число x, для которого $x - 5$, x и $x + 5$ являются квадратами.

Эту задачу ни в коем случае нельзя назвать простой — самое краткое ее решение таково:

$$x = \frac{1681}{144} = \left(\frac{41}{12}\right)^2.$$

В этом случае

$$x-5=\frac{961}{144}=\left(\frac{31}{12}\right)^2 \text{ и } x+5=\frac{2401}{144}=\left(\frac{49}{12}\right)^2.$$

Леонардо нашел решение и включил его в «Книгу квадратов». Он получил ответ при помощи общей формулы, связанной с формулой Евклида/Диофанта для пифагоровых троек. Из нее Леонардо получил три целых квадрата с общей разностью 720, а именно: 31^2, 41^2 и 49^2. Затем он разделил их на $12^2 = 144$, чтобы получить три квадрата с общей разностью 720/144, что равняется 5[38]. В терминах пифагоровых троек можно взять треугольник со сторонами 9, 40 и 41 и площадью 180 и разделить на 36. Получим треугольник со сторонами 20/3, 3/2, 41/6. Площадь его равняется 5.

Именно у Леонардо мы находим латинское слово *congruum* для обозначения набора из трех квадратов в арифметической прогрессии. Позже Эйлер пользовался словом *congruere*, «сходятся». Первые десять конгруэнтных чисел и соответствующие простейшие пифагоровы тройки приведены в табл. 3. Никаких простых закономерностей здесь не видно.

Таблица 3. Первые десять конгруэнтных чисел и соответствующие им пифагоровы тройки

d	**Пифагорова тройка**
5	3/2, 20/3, 41/6
6	3, 4, 5
7	24/5, 35/12, 337/60
13	780/323, 323/30, 106921/9690
14	8/3, 63/6, 65/6
15	15/2, 4, 17/2
20	3, 40/3, 41/3
21	7/2, 12, 25/2
22	33/35, 140/3, 4901/105
23	80155/20748, 41496/3485, 905141617/72306780

Первоначальным прогрессом в этом вопросе мы обязаны в первую очередь арабским математикам, показавшим, что числа 5, 6, 14, 15, 21, 30, 34, 65, 70, 110, 154 и 190, а также еще 18 больших чисел, являются конгруэнтными. Леонардо, Анджело Дженокки (1855) и Андре Жерарден (1915 г.) добавили к этим числам 7, 22, 41, 69, 77 и еще 43 числа, не превосходящих 1000. Леонардо в 1225 г. объявил, что число 1 не конгруэнтно, но не привел никаких доказательств. В 1569 г. Ферма доказал это. К 1915 г. все конгруэнтные числа меньше 100 были определены, но проблема плохо поддавалась решению, и еще в 1980 г. статус многих чисел меньше 1000 оставался неопределенным. О сложности проблемы можно судить по тому, как Л. Бастьен открыл конгруэнтность числа 101. Стороны соответствующего прямоугольного треугольника равны:

$$\begin{array}{c} 7110240645789550100000 \\ 118171431852779451900 \end{array},$$

$$\begin{array}{c} 3967272806033495003922 \\ 118171431852779451900 \end{array},$$

$$\begin{array}{c} 4030484925899520003922 \\ 118171431852779451900 \end{array}.$$

Он нашел эти числа в 1914 г. вручную. К 1986 г., когда считать благодаря компьютерам стало проще, Г. Крамарц нашел все конгруэнтные числа до 2000.

В какой-то момент было замечено, что другое, но связанное с этой задачей уравнение $y^2 = x^3 - d^2x$ имеет решение x, y в целых числах тогда и только тогда, когда d конгруэнтно[39]. В одном направлении это наблюдение очевидно: правая часть уравнения представляет собой произведение x, $x - d$ и $x + d$, а если все сомножители являются квадратами, то квадратом является и произведение. Обратное утверждение получить также несложно. Такая формулировка задачи сразу переводит ее в богатые и процветающие владения теории чисел. Для любого заданного d это уравнение задает y^2, равный кубическому многочлену от x, и таким образом определяет эллиптическую кривую. Так что проблема конгруэнтных чисел — частный случай вопроса, ответить на который мечтают многие специалисты по теории чисел: при каких условиях эллиптическая кривая содержит хотя бы одну рациональную точку? Вопрос этот далеко не очевиден, даже для только что упомянутого частного случая эллиптической кривой. К примеру, 157 — число конгруэнтное, но гипотенуза *простейшего* прямоугольного треугольника с такой площадью равна

$$\begin{array}{c} 224403517704336969924557513090674863160948472041 \\ 8912332268928859588025535178967163570016480830 \end{array}.$$

Прежде чем продолжить, мы позаимствуем у Леонардо его уловку — ту самую, что помогла перейти от 720 к 5, — и применим ее в самом общем виде. Умножив любое конгруэнтное число d на квадрат n^2 целого n, мы получим также конгруэнтное число. Чтобы убедиться в этом, достаточно взять любую рациональную пифагорову тройку, соответствующую треугольнику с площадью d, и умножить стороны на n. Площадь треугольника увеличится в n^2 раз. То же произойдет и при делении на n; площадь уменьшится в n^2 раз. Результат этого процесса будет целым только в том случае, если площадь делится нацело на квадрат целого числа (т. е. имеет квадратный делитель), так что при поиске конгруэнтных чисел достаточно работать только с числами, не имеющими такого делителя. Приведем первые несколько чисел, не имеющие квадратного делителя:

$$1, 2, 3, 5, 6, 7, 10, 11, 13, 14, 15, 17, 19.$$

Теперь можно сформулировать критерий Таннелла. Нечетное число d, не имеющее квадратных делителей, конгруэнтно тогда и только тогда, когда число (положительных или отрицательных) целых решений x, y, z уравнения

$$2x^2 + y^2 + 8z^2 = d$$

точно вдвое превосходит число решений уравнения

$$2x^2 + y^2 + 32z^2 = d.$$

Четное число d, не имеющее квадратных делителей, конгруэнтно тогда и только тогда, когда

$$8x^2 + 2y^2 + 16z^2 = d$$

точно вдвое превосходит число решений уравнения

$$8x^2 + 2y^2 + 64z^2 = d.$$

Эти результаты куда полезнее, чем может показаться на первый взгляд. Поскольку все коэффициенты уравнения положительны, x, y и z по модулю не могут превосходить некие числа, кратные корню квадратному из d. Из этого следует, что число решений конечно и их можно найти систематическим поиском с применением некоторых полезных уловок. Приведем полный расчет нескольких примеров с небольшими d:

- Если $d = 1$, то единственными решениями первого уравнения являются $x = 0$, $y = \pm 1$, $z = 0$. То же относится и ко второму уравнению. Так что оба уравнения имеют по два решения, и, следовательно, критерий не выполняется.
- Если $d = 2$, то единственными решениями первого уравнения являются $x = \pm 1$, $y = 0$, $z = 0$. То же относится и ко второму уравнению. Так что оба уравнения имеют по два решения, и, следовательно, критерий не выполняется.
- Если $d = 3$, то единственными решениями первого уравнения являются $x = \pm 1$, $y = \pm 1$, $z = 0$. То же относится и ко второму уравнению. Так что оба уравнения имеют

по четыре решения, и, следовательно, критерий не выполняется.

- Если $d = 5$ или 7, то первое уравнение не имеет решений. То же относится и ко второму уравнению. Поскольку дважды нуль равняется нулю, критерий выполняется.
- Если $d = 6$, то мы должны использовать критерий для четных чисел. Здесь опять же оба уравнения не имеют решений, и критерий выполняется.

Эти простые расчеты показывают, что 1, 2, 3, 4 ($= 2^2 \times 1$) не являются конгруэнтными, а 5, 6 и 7 — являются. Анализ несложно продолжить, и в 2009 г. команда математиков применила тест Таннелла ко всем числам до триллиона, обнаружив при этом ровно 3 148 379 694 конгруэнтных числа. Исследователи проверили результат, повторив все расчеты дважды на разных компьютерах с использованием разных алгоритмов и программ, написанных двумя независимыми группами программистов. Билл Харт и Гонсало Торнариа пользовались компьютером Selmer в Уорикском университете. Марк Уоткинс, Дэвид Харви и Роберт Брэдшоу работали с компьютером Sage в Вашингтонском университете.

Однако во всех этих расчетах есть пробел. Таннелл доказал, что, если число d конгруэнтно, оно должно удовлетворять его критерию. Таким образом, если критерий не выполняется, число не конгруэнтно. Однако он не сумел доказать обратного: если число удовлетворяет его критерию, то оно обязательно конгруэнтно. Именно это необходимо нам, чтобы сделать вывод о конгруэнтности чисел 5, 6 и 7. В данных конкретных случаях мы можем найти подходящие пифагоровы тройки, но в общем случае это нам не поможет. Таннелл сумел показать, что обратное утверждение, о котором идет речь, непосредственно следует из гипотезы Берча–Свиннертон-Дайера, но она тоже пока не доказана.

Гипотезу Берча–Свиннертон-Дайера, как и несколько других задач тысячелетия, сложно даже сформулировать. (А вы думали, что можно получить миллион долларов, сделав что-нибудь простое?) Однако настойчивость всегда окупается, ведь в процессе работы мы осознаем глубину и оцениваем давние исторические традиции теории чисел. Если вы внимательно посмотрите на название гипотезы, то заметите, что одно тире в нем длиннее другого. Дело в том, что эту гипотезу выдвинули не математики Берч, Свиннертон и Дайер, а Брайан Берч и Питер Свиннертон-Дайер. Ее полная формально-математическая формулировка сложна для непосвященных, но речь в ней идет о фундаментальном вопросе диофантовых уравнений — алгебраических уравнений, решения которых ищутся в целых или рациональных числах. Вопрос этот предельно прост: при каких условиях эти уравнения имеют решения?

В главе 6, где речь шла о гипотезе Морделла, и в главе 7, посвященной Великой теореме Ферма, мы встретились с одним из чудеснейших инструментов математики — эллиптическими кривыми. Морделл в свое время высказал, как тогда казалось, случайную догадку, предположив, что число рациональных решений алгебраического уравнения с двумя переменными зависит от топологии соответствующей комплексной кривой. Если род равен 0 — кривая топологически представляет собой сферу, — решения задаются формулой. Если род равен 1 — кривая топологически представляет собой тор, т. е. является эллиптической кривой, — то все рациональные решения могут быть построены из подходящего конечного списка путем приложения структуры группы. Если род равен 2 или больше — кривая топологически представляет собой тор с g отверстиями, где $g \geq 2$, — то число решений конечно. Как мы уже видели, Фальтингс доказал эту замечательную теорему в 1983 г.

Рациональные решения уравнений эллиптических кривых обладают одним поразительным свойством: благодаря геометрической конструкции, показанной на рис. 28 в главе 6, они образуют группу. Получившаяся структура называется группой Морделла–Вейля, и специалисты по теории чисел очень хотели бы иметь возможность вычислять ее. Для этого нужно найти систему генераторов: рациональных решений, из которых при помощи оператора группы могут быть получены все остальные. Если это не удается, то хотелось бы по меньшей мере определить основные характеристики группы, хотя бы ее величину. Здесь, однако, многое еще непонятно. Иногда группа бесконечна и порождает бесконечно много рациональных решений, иногда конечна, и тогда число рациональных решений тоже конечно. Было бы полезно иметь возможность определить, к какой категории относится конкретный случай. Но что нам по-настоящему хотелось бы знать, так это абстрактную структуру группы.

Доказательство Морделла, что конечный список генерирует все решения, говорит о том, что группа должна состоять из конечной группы и решетчатой группы. Решетчатая группа включает в себя все списки целых чисел конкретной конечной длины. Если длина чисел, к примеру, три, то группа состоит из всех списков (m_1, m_2, m_3) целых чисел, и эти списки складываются очевидным образом:

$$(m_1, m_2, m_3) + (n_1, n_2, n_3) = (m_1 + n_1, m_2 + n_2, m_3 + n_3).$$

Длина списка называется рангом группы (и геометрически представляет собой размерность решетки). Если ранг группы 0, группа конечна. Если ранг не равен нулю, группа бесконечна. Поэтому, чтобы понять, сколько существует решений, нам необязательно знать полную структуру группы. Достаточно знать ее ранг. Именно об этом говорит гипотеза Берча–Свиннертон-Дайера.

В 1960-е гг., когда компьютеры только-только входили в нашу жизнь, одна из первых таких машин появилась в Кембриджском университете. Называлась она EDSAC, что означало «электронно-счетная машина с запоминающим устройством на линиях задержки». Название показывает, как гордились создатели этой машины устройством ее памяти, посылавшей звуковые волны по трубкам с ртутью и затем направлявшей их вновь к началу. Размером этот компьютер был с большой грузовик. Я хорошо помню, как в 1963 г. мне устроили экскурсию по нему. Цепи компьютера были сделаны на основе тысяч ключей — электронных ламп. Вдоль всех стен стояли широкие стеллажи с запасными лампами, которые то и дело надо было менять — так часто они сгорали.

Питера Свиннертон-Дайера эллиптические кривые интересовали с диофантовой стороны: в первую очередь ему хотелось понять, сколько существовало бы решений, если заменить кривую ее аналогом на конечном поле с простым числом p элементов. Иными словами, ему хотелось изучить применявшуюся Гауссом уловку с работой «по модулю p». При помощи компьютера он вычислял эти числа для большого числа простых и искал среди них интересные закономерности.

Постепенно у него появились определенные подозрения. Его научный консультант Джон Кассельс испытывал сильные сомнения, но по мере появления все новых и новых данных он тоже поверил, что в этой идее что-то есть. Компьютерные эксперименты, проведенные Свиннертон-Дайером, указывали вот на что. У специалистов по теории чисел есть стандартный метод записи любого уравнения в целых числах по определенному модулю — вспомните модулярную арифметику или «арифметику часов» по модулю 12 в главе 2. Поскольку все законы алгебры приложимы в этом варианте арифметики, любое решение первоначального уравнения становится и решением «урезанного» уравнения по этому модулю. Все задействованные числа образуют конечный список — к примеру, для арифметики часов в этом списке всего 12 чисел, — поэтому все решения можно найти методом проб и ошибок. В частности, для каждого заданного модуля можно подсчитать, сколько существует решений. Кроме того, решения по каждому модулю накладывают определенные ограничения на решения первоначального уравнения и иногда даже помогают доказать, что такие решения существуют. Поэтому у специалистов по теории чисел выработался рефлекс рассматривать варианты любого уравнения по разным модулям, и простые числа особенно полезны в качестве таковых.

Таким образом, чтобы выяснить что-нибудь полезное об эллиптической кривой, можно рассмотреть все простые числа до определенного предела. Для каждого простого числа можно определить, сколько точек лежит на кривой по модулю этого числа. Берч заметил, что компьютерные эксперименты Свиннертон-Дайера показывают интересную закономерность, если разделить число таких точек на простое число, по модулю которого все рассматривалось. Затем следует перемножить результаты такого деления для всех простых чисел до заданного предела включительно и отложить результаты для последовательных простых чисел на логарифмической бумаге. Интересно, что все данные ложатся недалеко от прямой линии, крутизна которой представляет собой ранг данной эллиптической кривой. Это позволяло предложить гипотетическую формулу для числа решений, связанных с любым простым модулем[40].

Источник этой формулы, однако, не теория чисел: в ней задействован комплексный анализ, очень любимый в XIX в. и, по счастливому стечению обстоятельств, гораздо более элегантный, чем старомодный действительный анализ. В главе 9, посвященной гипотезе Римана, мы видели, как анализ вытягивает свои щупальца во всех направлениях и проникает в близкие и не очень области математики. Особенно удивительные и мощные связи возникли у него с теорией чисел. Формула Свиннертон-Дайера позволила выдвинуть более подробную гипотезу о типе комплексной функции (я упоминал ее в главе 9), известной как L-функция Дирихле. Эта функция аналогична для эллиптических кривых известной дзета-функции Римана. Эти два математика, очевидно, пытались обогнать время — ведь тогда не было даже наверняка известно, что у каждой эллиптической кривой *есть* L-функция Дирихле. Это было достаточно произвольное предположение, в пользу которого почти не было данных, но чем дальше шло развитие, тем правдоподобнее казалось это предположение. Это был не прыжок в неведомое, а изумительно точное и дальновидное проявление утонченной математической интуиции. Вместо того чтобы подняться на плечах гигантов, как чаще всего бывает в науке, Берч и Свиннертон-Дайер поднялись на собственных плечах — они были способны самостоятельно держаться в воздухе.

Основной инструмент комплексного анализа — выражение функции в виде степенного ряда, похожего на многочлен, но содержащего бесконечно много слагаемых с все более и более высокими степенями переменной, которую в этой области традиционно обозначают s. Чтобы выяснить, что функция делает около какой-то конкретной точки, скажем, 1, следует использовать степени $(s - 1)$. Гипотеза Берча–Свиннертон-Дайера утверждает, что если разложение L-функции Дирихле в степенной ряд возле 1 выглядит как

$$L\ (C, s) = c\ (s - 1)^r + \text{слагаемые более высоких степеней},$$

где c — ненулевая константа, то ранг кривой равен r, и наоборот. На языке комплексного анализа это утверждение принимает вид: $L\ (C, s)$ имеет в точке $s = 1$ нуль r-го порядка.

Главное здесь — не точное выражение, о котором идет речь; главное — то, что для любой заданной эллиптической кривой существует аналитическая формула с использованием соответствующей комплексной функции, при помощи которой можно точно узнать, сколько независимых рациональных решений необходимо найти, чтобы определить их все.

Возможно, простейший способ продемонстрировать, что гипотеза Берча–Свиннертон-Дайера имеет смысл и значение, — это упомянуть о том, что максимальный известный ранг равен 28. Иными словами, существует эллиптическая кривая с набором из 29 рациональных решений, позволяющим получить все остальные рациональные решения. Более того, меньшего набора рациональных решений, который позволял бы это сделать, не существует. Хотя известно, что кривые такого ранга существуют, конкретных примеров до сих пор не найдено. Максимальный ранг, для которого имеется конкретный пример, равен 18. Соответствующую кривую нашел в 2006 г. Ноам Элкис, и выглядит она так:

$$\begin{gathered} y^2 + xy = x^3 - 2617596009270588409631170178770120390355 \\ 6438969515x + 5106938147613148648974217710037377208 9 \\ 7791032538905678483 26. \end{gathered}$$

Я привел нестандартный вид «y^2 = кубический многочлен от x», но данную запись можно привести к стандартному виду за счет дополнительного увеличения коэффициентов. Считается, что ранг может быть сколь угодно большим, но это до сих пор не доказано. Если судить по уже имеющимся данным, ранг не может быть больше некоего фиксированного числа.

Большая часть утверждений, которые мы можем доказать, относится к кривым рангов 0 и 1. Если ранг равен 0, то существует конечное число рациональных решений. Если ранг равен 1, то одно конкретное решение позволяет получить почти все остальные, за исключением, возможно, конечного числа решений. Эти два случая включают все эллиптические кривые вида $y^2 = x^3 + px$, где p — простое число вида $8k + 5$ (такое, как 13, 29, 37 и т. д.). Предполагается, что в этих случаях ранг всегда равен 1, что подразумевает существование бесконечного числа рациональных решений. Эндрю Бремнер и Кассельс доказали верность этого утверждения для всех простых чисел соответствующего вида до 1000. Оказалось, что, даже если ранг известен и мал, найти решения, позволяющие получить почти все остальные, может быть очень трудно. Они выяснили, что для $p = 877$ *простейшим* решением такого рода является рациональное число

$$\begin{gathered} 37549452812716293105504069942092792346201 \\ 62159877768715054254632207806972380441 00 \end{gathered}.$$

Доказано огромное число теорем, имеющих отношение к гипотезе Берча–Свиннертон-Дайера (обычно с очень серьезными формальными ограничениями), но это пока мало помогло в продвижении к полному решению этой задачи. В 1976 г. Коутс и Уайлс обнаружили первые указания на то, что эта гипотеза может быть верна. Они доказали, что один частный случай эллиптической кривой имеет ранг 0, если *L*-функция Дирихле не обращается в нуль в точке 1. Для такой эллиптической кривой число решений связанного с ней диофантова уравнения конечно, возможно, равно нулю, и определить это можно по соответствующей *L*-функции. После этого момента удалось сделать несколько технических шагов, по-прежнему ограниченных в основном рангами 0 и 1. В 1990 г. Виктор Колывагин доказал, что гипотеза Берча–Свиннертон-Дайера верна для рангов 0 и 1.

Более детальные гипотезы, требующие серьезной компьютерной поддержки, соотносят константу *c* в гипотезе Берча–Свиннертон-Дайера с различными концепциями теории чисел. Существуют аналогичные гипотезы — впрочем, столь же загадочные, — для алгебраических числовых полей. Известно также, что большинство (в точном смысле) эллиптических кривых имеет ранг 0 или 1. В 2010 г. Манджул Бхаргава и Арул Шанкар объявили, что им удалось доказать: средний ранг эллиптической кривой не превосходит 7/16. Если это доказательство и доказательство некоторых других недавно опубликованных теорем будут признаны математическим сообществом, то получится, что гипотеза Берча–Свиннертон-Дайера верна для ненулевой доли всех эллиптических кривых. Однако речь пока идет о простейших кривых, не представляющих, по существу, кривые более сложной структуры, ранга 2 и более. Они пока остаются для нас загадкой.

15

Комплексные циклы

Гипотеза Ходжа

Некоторые области математики вполне можно соотнести с тем, с чем мы встречаемся в повседневной жизни. Уравнение Навье–Стокса невозможно встретить на кухне, но мы все понимаем, что такое жидкости, и представляем, как они текут. Другие области можно соотнести с эзотерическими вопросами переднего края науки: так, чтобы разобраться в квантовой теории поля, нужна хотя бы докторская степень в области математической физики, но аналогии с электричеством и магнетизмом или такие хоть сколько-то представимые образы, как волны вероятности, позволяют кое-что понять. Третьи можно объяснить при помощи картинок, и хороший пример тому — гипотеза Пуанкаре. Но некоторые области математики не поддаются ни одному из перечисленных способов и никак не позволяют сделать сложные абстрактные понятия доступными.

Гипотеза Ходжа, сформулированная в 1950 г. шотландским геометром Уильямом Ходжем, — одна из таких задач. Проблемы здесь возникают не из-за доказательства, поскольку его просто нет. Все дело в утверждении. Вот так примерно эта задача сформулирована на сайте Института Клэя:

> «На любом невырожденном проективном комплексном алгебраическом многообразии любой класс Ходжа пред-

> ставляет собой рациональную линейную комбинацию классов алгебраических циклов».

На первый взгляд в этой формулировке понятны, пожалуй, только предлоги и такие слова, как «любой». Остальное понятно, как отдельные слова: «многообразие», «класс», «рациональный», «цикл». Но образы, порождаемые этими словами, — виды в живой природе, школа, разум без эмоций, какой-то повторяющийся процесс — явно относятся не к тому, что имел в виду Институт Клэя. Остальное еще более очевидный жаргон. Но не просто жаргон ради жаргона — не сложные слова, за которыми прячется профессиональная лексика. Точнее, это простые слова для обозначения сложных вещей. В обычном языке нет готовых названий для подобных концепций, так что часть приходится заимствовать в других областях, а часть изобретать заново.

Если говорить о хорошем, то здесь у нас появляются немалые возможности. Можно сказать, что гипотеза Ходжа лучше представляет реальную математику XX и XXI вв., чем любая другая из рассмотренных в этой книге тем. Подойдя к ней надлежащим образом, мы сможем получить представление о том, насколько концептуально продвинута на самом деле современная передовая математика. В сравнении со школьной математикой она как Эверест в сравнении с кучкой земли, оставленной кротом.

Но, может быть, это всего лишь пустое сотрясание воздуха, претенциозная чепуха, которой занимаются отшельники в башнях из слоновой кости? Если ни один нормальный человек не в состоянии понять, о чем идет речь, зачем впустую переводить деньги налогоплательщиков на тех, кто думает о подобных вещах? Однако давайте взглянем на это с другой стороны. Предположим, любой человек мог бы понять все, о чем думают математики. Неужели тогда вы с удовольствием отдали бы математикам деньги налогоплательщиков? Разве им платят не за профессиональные знания? Если бы все было настолько просто и понятно, что разобраться в этом мог бы любой, зачем вообще надо было бы готовить математиков? А если бы каждый умел налаживать центральное отопление и сваривать трубы, для чего были бы нужны водопроводчики?

Я не могу сказать вам, как именно могла бы быть с пользой применена гипотеза Ходжа. Но я могу объяснить, насколько важное место она занимает в математике. Современная математика — единый организм, так что значительное продвижение в любой из основных областей со временем принесет вполне материальный доход, измеряемый в долларах и центах. Может быть, сегодня мы не найдем на своей кухне ни одного прибора, сделанного на основе этой гипотезы, но завтра — кто знает? Тесно связанные с ней математические концепции уже доказывают свою полезность в различных областях науки — от квантовой физики и теории струн до робототехники.

Иногда новые математические идеи получают практическое применение почти сразу. Иногда этот процесс занимает не одно столетие. Быть может, в последнем случае лучше было бы подождать, пока возникнет нужда в этих идеях, а затем ударными темпами провести их разработку? Быть может, все математические задачи, не имеющие немедленного и очевидного применения, следует откладывать в дальний ящик на будущее? Однако если бы мы так поступали, то всегда отставали бы от жизни, поскольку математики уже несколько сотен лет играют в догонялки с прикладной наукой. Да и не всегда можно точно сказать, какая идея необходима в данный момент. Как вы думаете, понравилось бы вам, если бы никто даже не задумался о производстве кирпичей, пока вы не пригласили бы рабочих для строительства дома? Чем оригинальнее математическая концепция, тем более маловероятно, что она родится в результате ударной разработки.

Куда разумнее было бы позволить математической науке развиваться по собственным законам и не ждать от нее немедленной пользы. Не пытайтесь выбирать лучшее, позвольте ей расти свободно. Математики стоят недорого: им, в отличие от физиков-экспериментаторов, не нужно дорогостоящее оборудование (на Большой адронный коллайдер уже потрачено €7,5 млрд, и расходы растут). Кроме того, в качестве компенсации математики обучают студентов. И вряд ли было бы разумно не разрешить некоторым из них работать над гипотезой Ходжа, если эта проблема их захватила.

Я планирую разобрать приведенную формулировку гипотезы Ходжа слово за словом. Простейшая из встречающихся в ней концепций — «алгебраическое многообразие». Это естественное следствие декартова подхода, когда тот при помощи координатной сетки связал геометрию с алгеброй (см. главу 3). При этом крохотный набор инструментов-кривых, введенный Евклидом и его последователями, — прямая, окружность, эллипс, парабола, гипербола — превратился в бездонный рог изобилия. Прямая линия — основа евклидовой геометрии — представляет собой совокупность точек, удовлетворяющих соответствующему алгебраическому уравнению: к примеру, $y = 3x + 1$. Замените тройку и единицу на другие числа — и получите другие прямые линии. Окружности нуждаются в квадратных уравнениях — как и эллипсы, параболы и гиперболы. В принципе, все, что можно определить геометрически, можно интерпретировать и иначе — алгебраически, — и наоборот. Так что, система координат делает геометрию ненужной? Или, может быть, она делает ненужной алгебру? Зачем пользоваться двумя инструментами, если оба они делают одно и то же?

У меня в гараже в ящике с инструментами есть и молоток, и клещи. Дело молотка — забивать в дерево гвозди. Дело клещей — вытаскивать их оттуда. Хотя, в принципе, гвозди можно забить и клещами, а у молотка с обратной стороны есть раздвоенный конец, предназначенный специально для выдергивания гвоздей. Зачем же мне оба инструмента? Затем, что одни вещи лучше делать молотком, а другие — клещами. Так же обстоит дело с алгеброй и геометрией: одни подходы более естественно реализуются при помощи геометрии, другие — при помощи алгебры. Главное — связь между ними. Если алгебраическое мышление буксует, переключайтесь на геометрию.

Координатная геометрия предлагает новую свободу выдумывать кривые. Просто напишите уравнение — и смотрите на его решения. Если ваше уравнение не слишком глупое, вроде $x = x$, должна получиться кривая. (Решениями уравнения $x = x$ является вся координатная плоскость.) К примеру, я мог бы записать уравнение $x^3 + y^3 = 3xy$, решения которого можно увидеть на рис. 45. Эта кривая — декартов лист, и вы не найдете ее у Евклида. Ассортимент новых кривых, которые может выдумывать каждый, буквально бесконечен.

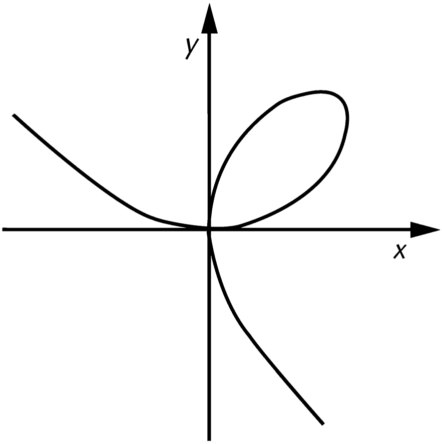


Рис. 45. Декартов лист

Математики всегда стремятся к обобщениям — это рефлекс, он включается автоматически. Стоит кому-нибудь натолкнуться на интересную идею, и тут же все задаются вопросом: возможно ли что-нибудь подобное в более общем случае? Идея Декарта, в частности, имеет по крайней мере три серьезных варианта обобщения, или модификации, и все они необходимы для понимании гипотезы Ходжа.

Во-первых, что происходит, если мы работаем с пространствами, отличными от плоскости? Трехмерное евклидово пространство имеет три координаты (x, y, z) вместо двух. В пространстве одно уравнение, как правило, определяет поверхность, а два уравнения — кривую, по которой поверхности пересекаются. Три уравнения, как правило, определяют точку. (Говоря «как правило», я имею в виду, что бывают и исключения, но они очень необычны и удовлетворяют особым условиям. Что-то подобное мы видели на плоскости в случае того самого «глупого» уравнения $x = x$.)

Здесь мы опять же можем придумывать новые уравнения и тем самым определять новые поверхности или кривые, которых нет у Евклида. В XIX в. это было модным занятием. Можно было даже опубликовать статью про новую поверхность, если у вас было что сказать о ней — что-нибудь по-настоящему интересное. В качестве типичного примера можно вспомнить поверхность, представленную Куммером в 1864 г., с уравнением

$$x^4 + y^4 + x^4 - y^2z^2 - z^2x^2 - x^2y^2 - x^2 - y^2 - z^2 + 1 = 0.$$

На рис. 46 представлен соответствующий график. Самое интересное в нем — 16 «двойных точек», где поверхность напоминает поверхности двух конусов, соединенных вершина к вершине. Как раз 16 — максимально возможное число таких точек для поверхности четвертого порядка, т. е. поверхности, описываемой уравнением четвертой степени. Это обстоятельство показалось достаточно интересным для публикации.

К XIX в. математики успели познать пьянящие радости пространств высоких измерений. Нет нужды останавливаться на трех координатах; почему не попробовать четыре, пять, шесть… миллион? И это не пустопорожние рассуждения. Это алгебра множества уравнений с множеством переменных, а они всплывают то и дело в самых разных точках математического ландшафта. К примеру, они упоминались в главе 5 (гипотеза

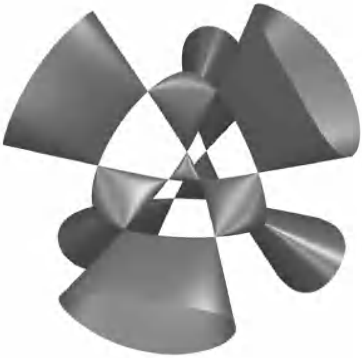


Рис. 46. Поверхность четвертого порядка Куммера с 16 двойными точками

Кеплера) и главе 8 (задача трех тел). Не идет речь и о пустых искусственных обобщениях: возможность размышлять о подобных вещах не только алгебраически, но и геометрически — мощный инструмент, который нет смысла ограничивать двумя или тремя измерениями просто потому, что только в таких пространствах мы можем рисовать картинки и строить модели.

Слово «измерение» может казаться внушительным и загадочным, но в данном контексте его значение вполне прозрачно: сколько вам нужно координат. К примеру, в четырехмерном пространстве четыре координаты (x, y, z, w), и в математическом смысле этого достаточно для определения. В четырех измерениях единственное уравнение обычно определяет трехмерную «гиперповерхность», два уравнения — поверхность (два измерения), три уравнения — кривую (одно измерение), а четыре — точку (нуль измерений). Каждое новое уравнение расправляется с одним измерением, т. е. с одной переменной. Так что мы можем предсказать, что в пространстве 17 измерений 11 уравнений определяют шестимерный объект, за исключением редких (и легко опознаваемых) случаев, когда некоторые из уравнений избыточны.

Объект, определенный таким образом, называется алгебраическим многообразием. В русском языке слово «многообразие» употребляется и в топологии, и в дифференциальной геометрии (топологии пополам с дифференциальным исчислением), и в алгебраической геометрии. В некоторых других языках традиционно существует два различных термина (в частности, в английском языке используются слова manifold и variety)[41]. Конечно, алгебраическое многообразие можно было бы называть «многомерным пространством, определенным системой алгебраических уравнений», но вы сами, вероятно, понимаете, почему так никто не говорит.

Второй многообещающий способ обобщить представления координатной геометрии состоит в том, чтобы разрешить комплексные координаты. Припомним, кстати, что в системе комплексных чисел существует число нового типа i, квадрат которого равен –1. Зачем усложнять все на свете таким странным образом? Затем, что алгебраические уравнения на множестве комплексных чисел ведут себя гораздо лучше. На множестве действительных чисел квадратное уравнение может иметь два решения или ни одного. (Оно может также иметь одно решение, но в определенном — и весьма разумном — смысле лучше считать, что одно решение повторяется дважды.) На множестве комплексных чисел квадратное уравнение *всегда* имеет два решения (опять же если корректно учитывать повторяющиеся решения). В некоторых случаях такое свойство может оказаться очень полезным. Можно сказать: «Решаем уравнение для седьмой переменной» — и быть уверенным, что такое решение действительно существует.

Тем не менее, хотя в этом отношении все очень удобно, некоторые свойства комплексной алгебраической геометрии без привычки воспринимаются довольно тяжело. Если говорить о действительных переменных, то там прямая может пересекать окружность в двух точках, касаться ее или проходить в стороне и не иметь с ней общих точек. В случае комплексных переменных третья возможность исчезает. Но если привыкнуть к изменениям, то окажется, что комплексные алгебраические многообразия ведут себя куда лучше, чем действительные. Иногда действительные переменные необходимы, но в большинстве случаев в комплексном контексте работать удобнее. Во всяком случае нам теперь известно, что представляет собой комплексное алгебраическое многообразие.

Как насчет слова «проективное»? Это третье обобщение, и для него требуется несколько иное представление о пространстве. Проективная геометрия выросла из интереса, который живописцы эпохи Возрождения питали к законам перспективы, и в ней отсутствует особое поведение параллельных прямых. В евклидовой геометрии две прямые либо пересекаются, либо параллельны, и тогда они не встретятся никогда, сколько их ни продолжай. А теперь вообразите себя стоящим с кистью в руке перед мольбертом на бесконечной плоскости. Все готово, палитра ждет, а перед вами две параллельные прямые уходят к закатному горизонту, как два бесконечных идеально прямых железнодорожных рельса. Что вы видите и, соответственно, что появится на вашем холсте? Вовсе не две линии, которые никак не могут сойтись. Вы увидите, как линии постепенно сближаются и на горизонте сходятся в точку.

Какой части плоскости соответствует горизонт? Той части, где встречаются параллельные линии. Но такого места нет. Горизонт на вашей картине представляет собой границу изображения плоскости. Если с окружающим миром все в порядке, то горизонт должен быть изображением границы плоскости. Но у плоскости нет границ. Она продолжается бесконечно. Все это слегка сбивает с толку, как будто часть евклидовой плоскости куда-то пропала. «Проектируя» плоскость (ту самую, с рельсами) на другую плоскость (ваш холст на мольберте), вы получаете на картине линию — горизонт, — которая не является проекцией никакой линии на изображаемой плоскости.

Существует способ избавиться от этой загадочной аномалии: добавить к евклидовой плоскости так называемую линию бесконечности, представляющую отсутствующий горизонт. После этого все сильно упрощается. Две прямые всегда встречаются в точке; прежнее представление о параллельных прямых соответствует случаю, когда две прямые встречаются в бесконечности. Эту идею после надлежащего осмысления можно совершенно разумно перевести на язык математики. Результат такого перевода и называется проективной геометрией. Это очень элегантный предмет, и математики XVIII и XIX вв. его обожали. Со временем оказалось, что сказать им по этому вопросу больше нечего — все уже сказано, и в таком состоянии эта область пребывала до тех пор, пока математики XX в. не решили обобщить алгебраическую геометрию на многомерные пространства и использовать комплексные числа. В этот момент стало ясно, что с тем же успехом можно довести дело до логического конца и вместо действительных решений систем алгебраических уравнений в евклидовом пространстве изучать комплексные решения в проективном пространстве.

Позвольте мне суммировать сказанное. Проективное комплексное алгебраическое многообразие похоже на кривую, определенную алгебраическим уравнением, за исключением того, что:

- число уравнений и переменных может быть любым по нашему желанию (алгебраическое многообразие);
- переменные могут быть комплексными, а не действительными (комплексность);
- переменные могут принимать бесконечные значения разумным образом (проективность).

Добавим здесь же, что несложно разобраться и с еще одним термином из формулировки: с невырожденностью. Это слово означает, что многообразие является гладким и не имеет острых гребней или мест, где его форма сложнее, чем просто гладкий кусок пространства. Поверхность Куммера, например, имеет сингулярности в 16 двойных точках. Разумеется, нам нужно еще объяснить, что означает «гладкость», когда переменные комплексны и некоторые из них могут быть бесконечными, но на это есть рутинные общепринятые методики.

Вот мы и добрались почти до середины формулировки гипотезы Ходжа. Мы уже знаем, о чем идет речь, но пока не понимаем, как, по мнению Ходжа, эта штука должна себя вести. Теперь нам нужно разобраться с самыми глубокими и в то же время формальными аспектами: алгебраическими циклами, классами и особенно классами Ходжа. Однако самую суть я могу раскрыть прямо сейчас. Все это технические средства, помогающие получить частичный ответ на фундаментальнейший вопрос о нашей обобщенной кривой: какой она формы? Оставшаяся часть формулировки — «рациональная линейная комбинация» — говорит о том, как в соответствии с общими надеждами следует ответить на этот вопрос.

Смотрите, как далеко мы продвинулись. Мы уже понимаем, что примерно представляет собой гипотеза Ходжа. Она говорит о том, что форму любой обобщенной поверхности, задаваемой некими уравнениями, можно определить при помощи каких-то алгебраических манипуляций с вещами, известными как циклы. Я мог бы сказать об этом в самом начале главы, но тогда эта формулировка вряд ли объяснила бы много больше, чем официальная. Теперь же, когда мы знаем, что такое многообразие, все понемногу проясняется.

Кроме того, все начинает сильно напоминать топологию. «Определение формы путем алгебраических вычислений» поразительно похоже на идеи Пуанкаре об алгебраических инвариантах топологических пространств. Так что следующий шаг потребует обсуждения алгебраической топологии. В активе Пуанкаре значится открытие трех важных типов инвариантов, определенных в терминах трех концепций: гомотопии, гомологии и когомологии. Нас в данном случае интересует когомология — и конечно (кто бы мог подумать!), именно ее объяснить труднее всего.

Я думаю, пора приступать.

В трехмерном пространстве с действительными координатами пересечением сферы и плоскости (если они, конечно, вообще пересекаются) является окружность. Сфера — это алгебраическое многообразие; окружность — тоже алгебраическое многообразие и притом входит в состав сферы. Мы называем это *подмногообразием*. В более общем случае, если взять уравнения (с большим числом переменных, комплексные, проективные), определяющие некое многообразие, и добавить к ним еще несколько уравнений, то некоторые решения — те, что не удовлетворяют новым уравнениям, — как правило, теряются. Чем больше у нас уравнений, тем меньше становится многообразие. Расширенная система уравнений определяет некоторую часть первоначального многообразия, и эта часть сама по себе тоже является многообразием — это подмногообразие.

При подсчете количества решений полиномиального уравнения иногда бывает удобно учесть одну и ту же точку несколько раз. Можно сказать, что совокупность решений состоит из множества точек, за каждой из которых мы «закрепляем» число, соответствующее его кратности. Можно, к примеру, иметь решения 0, 1 и 2 с кратностью 3, 7 и 4 соответственно. Многочлен в этом случае будет $x^3(x-1)^7(x-2)^4$, если вам это интересно. Каждая из трех точек $x = 0$, 1 или 2 является (достаточно тривиальным) подмногообразием множества комплексных чисел. Поэтому решения этого полиномиального уравнения можно описать как список из трех подмногообразий с прикрепленным к каждому из них целым числом (вроде этикетки).

Алгебраический цикл выглядит примерно так же. Вместо отдельных точек мы можем использовать любой конечный список подмногообразий, присоединив к каждому из них числовую метку, не обязательно целую. Меткой может быть отрицательное целое число, рациональное число, действительное или даже комплексное число. По разным причинам в гипотезе Ходжа в качестве меток используются рациональные числа, о чем свидетельствует формулировка «рациональная линейная комбинация». К примеру, в качестве первоначального многообразия может выступать единичная сфера в 11-мерном пространстве; тогда список, о котором идет речь, мог бы выглядеть так:

- семимерная гиперсфера (задаваемая такими-то уравнениями) с меткой 22/7;
- тор (задаваемый такими-то уравнениями) с меткой –4/5;
- кривая (задаваемая такими-то уравнениями) с меткой 413/6.

Не пытайтесь это представить или, если очень захочется, нарисуйте картинку в стиле комикса: три бесформенные кляксы с надписями. Каждая такая картинка, каждый список представляет один алгебраический цикл.

К чему устраивать такой шум и изобретать подобные абстракции? К тому, что они отражают самые существенные аспекты первоначального алгебраического многообразия. Специалисты по алгебраической геометрии заимствуют методы у топологов.

В главе 10, где речь шла о гипотезе Пуанкаре, мы говорили о муравье, вселенной которого является поверхность. Как может муравей определить форму своей вселенной, если он не в состоянии отойти в сторонку и посмотреть? В частности, как он сможет отличить сферу от тора? Представленное в той главе решение предусматривало использование замкнутых кривых — топологических автобусных маршрутов. Муравей перемещает эти петли по всей поверхности, выясняет, что происходит, если поставить их одну за другой — концом к началу, и вычисляет алгебраический инвариант пространства, известный как его фундаментальная группа. Слово «инвариант» означает, что топологически эквивалентные пространства имеют одну и ту же фундаментальную группу. Если группы различны, то различны и пространства. Именно этот инвариант привел Пуанкаре к его гипотезе. Однако бедному муравью непросто проверить все возможные в его вселенной маршруты, и это замечание отражает реальные математические тонкости в расчетах фундаментальных групп. Существует и более практичный инвариант, Пуанкаре его тоже исследовал. Процесс перемещения петель по поверхности называется гомотопией; альтернативный вариант называется похоже, но иначе — гомологией.

Я покажу вам простейший, самый конкретный вариант гомологии. Топологи быстро развили этот вариант, оптимизировали и обобщили его, превратив в мощнейшую математическую машину, которая получила название «гомологическая алгебра». Этот простой вариант позволит вам лишь слегка почувствовать, как все это работает, но ведь нам ничего больше и не нужно.

Муравей начинает с того, что обследует свою вселенную и составляет карту. Подобно любому профессиональному топографу, он покрывает вселенную сетью треугольников. Главное при этом — чтобы ни в одном треугольнике не оказалось дырки в поверхности. Проще всего обеспечить это, вставляя каждый треугольник в виде резиновой заплатки, как при ремонте велосипедной камеры. При этом каждый треугольник будет иметь хорошо определенную внутренность, топологически эквивалентную внутренности любого обычного треугольника на плоскости. Топологи называют такую треугольную заплатку топологическим диском, поскольку она эквивалентна кругу. Чтобы убедиться в этом, взгляните на рис. 36 в главе 10, где треугольник постепенно модифицируется в круг. Подобную заплатку невозможно поставить поверх отверстия, потому что отверстие создает туннель, связывающий внутреннюю часть треугольника с его внешней частью. Чтобы перекрыть отверстие, заплатке придется выйти за пределы поверхности, а муравью запрещено делать это.

Итак, муравей провел *триангуляцию* своей вселенной. Условие про заплатку гарантирует, что, имея полный список треугольников и зная, какой треугольник с какими граничит, можно восстановить топологию поверхности, т. е. ее форму в смысле топологической эквивалентности. Если бы можно было поехать в «Икею» и купить Универсальный муравьиный набор надлежащим образом промаркированных треугольников, то мы могли бы, склеив аккуратно сторону A со стороной AA, сторону B со стороной BB и т. д., построить соответствующую поверхность. Сам муравей заперт на этой поверхности и потому не может построить ее модель, но он может быть уверен, что в принципе его карта содержит всю необходимую для построения информацию. Чтобы извлечь эту информацию, муравью придется проводить вычисления. При этом ему придется рассматривать уже не бесконечное число возможных петель, но все же достаточно большое их число: все замкнутые петли, проходящие вдоль ребер выбранной им сетки.

В гомотопии мы задаемся вопросом, можно ли сжать данную петлю непрерывно в точку. В гомологии мы задаемся другим вопросом: образует ли данная петля границу топологического диска? Иными словами, можно ли взять одну или несколько треугольных заплаток вместе таким образом, чтобы в сумме получился участок без отверстий с замкнутой границей?

На рис. 47 слева показана часть триангуляционной сети сферы — замкнутая петля и топологический диск, границей которого она является. Применив подходящие методики, можно доказать, что любая петля в триангуляционной сети сферы является такой границей: треугольные заплатки, а в более общем случае топологические диски, — это детекторы отверстий, а интуитивно понятно, что в сфере отверстий нет. Однако в торе отверстие имеется и в самом деле некоторые петли на торе не являются границами таких областей. На рис. 47 справа показана такая петля, проходящая сквозь центральное отверстие. Иными словами: просмотрев список петель и проверив, какие из них являются границами непрерывных областей, муравей может отличить сферическую вселенную от тороидальной.

Если наш муравей столь же умен, как Пуанкаре и другие топологи того времени, он сможет превратить эту идею в элегантный топологический инвариант — гомологическую группу

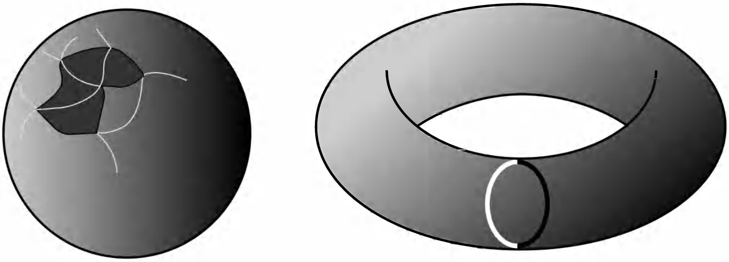


Рис. 47. Часть триангуляционной сети сферы: замкнутая петля (черные линии) и диск (темно-серая заливка), границей которого она является *(слева).* Петля на торе, не являющаяся границей диска (более светлая часть линии находится сзади) *(справа)*

своей поверхности. Базовая идея заключается в том, чтобы «сложить» две петли, нарисовав их обе. Однако то, что получилось, не является петлей, поэтому нам придется вернуться и начать заново. Более того, вернуться нам придется в самое начало, в те дни, когда мы только начинали свое знакомство с алгеброй. Моя учительница математики для начала поведала нам, что можно сложить количество яблок в одной кучке с количеством яблок в другой и получить общее количество яблок. Но нельзя сложить яблоки с апельсинами — разве что если хотите узнать общее число фруктов.

Сказанное верно в арифметике, хотя даже там приходится быть внимательным, чтобы не сосчитать одно и то же яблоко дважды, а в алгебре это уже неверно. Там вы можете складывать яблоки с апельсинами, не смешивая их. Более того, в высшей математике принято складывать вещи, которые никто в здравом уме и выдумывать-то не стал бы, не то что складывать. Свобода заниматься подобными вещами оказывается поразительно полезной и важной, и придумавшие их математики вовсе не были сумасшедшими, по крайней мере в этом отношении.

Для понимания некоторых идей, связанных с гипотезой Ходжа, мы должны иметь возможность складывать яблоки и апельсины, не записывая их все в обычные фрукты. Делать это на самом деле несложно. Сложно признать, что в этом занятии есть какой-то смысл. Многим из нас доводилось встречаться с подобными концептуальными блоками. Моя учительница рассказывала классу, что буквами обозначаются неизвестные числа, причем разные неизвестные числа обозначаются разными буквами. Если у вас есть a яблок и еще a яблок, то общее число яблок будет $a + a = 2a$. И это верно, сколько бы в реальности ни было яблок. Если вы возьмете $3a$ яблок и прибавите к ним еще $2a$ яблок, то всего получится $5a$, сколько бы в реальности ни было яблок. Сам символ, как и то, что он представляет, вовсе не имеет значения: если бы вместо $3a$ яблок у вас было $3b$ апельсинов, к которым вы прибавляли бы $2b$ апельсинов, то результат был бы $5b$[42]. Но что произойдет, если у вас будет $3a$ яблок и $2b$ апельсинов? Что будет, если сложить $3a$ и $2b$?

$$3a + 2b.$$

Вот и все. Эту сумму невозможно упростить и превратить в 5 чего-нибудь: по крайней мере нельзя без некоторых манипуляций с новой категорией — фруктами — и каких-то новых уравнений. Это лучшее, что можно получить: удовлетворитесь этим. Однако, начав с этого, вы вскоре сможете производить такие действия, как:

$$(3a + 2b) + (5a - b) = 8a + b$$

без всяких дополнительных рассуждений. И без новых видов фруктов.

Есть, правда, кое-какие оговорки. Я уже отметил, что при складывании яблока и яблока два яблока получится только в том случае, если первое яблоко не идентично второму. То же можно сказать и о более сложных комбинациях яблок и апельсинов. В алгебре считается, что для целей сложения все яблоки, о которых идет речь, различны между собой. Вообще-то принять такое условие часто имеет смысл даже в тех случаях, когда два яблока — или что мы там складываем — на самом деле могут оказаться идентичными. Одно яблоко плюс еще раз то же самое яблоко будет яблоко с кратностью два.

Привыкнув к этой идее, вы сможете пользоваться ею везде. Одна свинья плюс та же свинья получается свинья с кратностью два: свинья + свинья = 2 свиньи, что бы ни скрывалось на самом деле под словом «свинья». Свинья плюс корова будет свинья + корова. Треугольник плюс три круга будет треугольник + три круга. Суперпуперсфера плюс три гиперэллиптических квазикучи будет

суперпуперсфера + три гиперэллиптических квазикучи,

что бы все эти специальные термины ни означали (в данном случае ничего).

Можно даже разрешить отрицательные числа и говорить о вычитании 11 коров из трех свиней: 3 свиньи — 11 коров. Я понятия не имею, что представляют собой минус 11 коров, но я могу быть уверен, что если я прибавлю к ним шесть коров, то получу минус пять коров[43]. Это формальная игра с символами, и никакая реалистичная интерпретация здесь не требуется, не нужна или — зачастую — невозможна. Можно разрешить действительные числа: π свиней минус 2 коров. Комплексные числа. Любые сколь угодно причудливые числа, которые взбредут в голову математику. Этой идее можно придать чуть больше лоска и респектабельности, если рассматривать числа как *бирки,* навешенные свиньям и коровам. Тогда π свиней минус 2 коров можно рассматривать как свинью с биркой π рядом с коровой с биркой $- \quad 2$. Арифметика здесь применяется к биркам, а не к животным.

В гипотезе Ходжа тоже фигурирует подобная конструкция с дополнительными рюшечками и украшениями. Вместо животных в ней используются кривые, поверхности и их многомерные аналоги. Может показаться странным, но в результате получается не просто абстрактная чепуха, а глубокая связь между топологией, алгеброй, геометрией и анализом.

Чтобы привести в порядок математический аппарат гомологии, нам потребуется складывать петли, но не так, как мы делали это в фундаментальной группе, а так, как учила меня в свое время учительница. Мы будем просто записывать петли и ставить знак «+» между ними. Чтобы это имело смысл, мы будем работать не с отдельными петлями, а с конечными их наборами. Мы обозначим каждую петлю целым числом, которое будет соответствовать частоте встречаемости этой петли, и назовем такой набор *циклом*. Теперь наш муравей получает возможность складывать циклы. Для этого он должен объединить петли и сложить значения соответствующих маркеров. Результатом будет новый цикл. Возможно, рассказывая в главе 10 о путешествиях муравья, мне следовало взять мотоциклы, а не автобусы.

Когда мы занимались строительством фундаментальной группы, где «сложение» означает соединение петель концом к концу, там была одна техническая проблема. Добавление тривиальной петли к любой другой давало в результате *не совсем* ту же самую петлю, так что нулевая петля вела себя неправильно. Сложение прямой и обратной петель давало *не совсем* нулевую петлю, так что инверсия тоже работала некорректно. Чтобы решить эту проблему, решено было считать петли одинаковыми, если одну из них можно плавно преобразовать во вторую.

Для гомологии это вообще не проблема. Существует нулевой цикл (все маркеры нулевые), и для каждого цикла существует обратный к нему цикл (чтобы получить его, достаточно поменять знак у маркера цикла), поэтому мы имеем группу. Проблема в том, что это не та группа. Она ничего не говорит нам о топологии пространства. Чтобы разобраться в этом, мы воспользуемся аналогичной уловкой и более свободным подходом к тому, что считать нулем. Муравей режет пространство на треугольные заплатки, и граница каждой заплатки топологически достаточно тривиальна: ее можно свести в точку, просто сужая со всех сторон к середине. Таким образом, все граничные циклы должны быть эквивалентны нулевому циклу. Этот логический ход немного напоминает переход от обычных чисел к значениям по модулю (скажем, по модулю 12); мы делаем вид, что число 12 не имеет значения, и его можно назвать нулем. Здесь мы переводим циклы в плоскость гомологии, делая вид, что любые граничные циклы значения не имеют.

Следствия такой позиции очень серьезны. Теперь на алгебру циклов влияет топология пространства. Группа циклов по модулю границ является полезным топологическим инвариантом — гомологической группой поверхности. На первый взгляд этот инвариант зависит от того, какой вариант триангуляции выберет муравей, но если говорить об эйлеровой характеристике, то различные варианты триангуляции одной и той же поверхности приводят к одной и той же гомологической группе. Таким образом, муравей придумал алгебраический инвариант, при помощи которого можно различать поверхности. Искать его — довольно трудоемкое занятие, но хорошие инварианты невозможно получить без труда. Данный инвариант настолько эффективен, что с его помощью можно отличить не только сферу от тора, но тор с двумя отверстиями от тора с пятью отверстиями или с любым другим их количеством.

Гомология может показаться слишком сложной, но именно она положила начало целой серии топологических инвариантов. Кроме того, она основана на простых геометрических идеях: петлях, границах, объединении наборов, арифметических действиях с маркерами. Учитывая, что бедный муравей заперт на своей поверхности, просто поразительно, что он может узнать кое-что о своей вселенной при помощи разделения поверхности на треугольные кусочки, составления карты и некоторых алгебраических операций.

Можно естественным образом распространить гомологию на высшие измерения. Трехмерный аналог треугольника — тетраэдр; у него четыре вершины, шесть ребер, четыре треугольные грани и одна трехмерная «грань», его внутренность. В более общем случае в n измерениях можно определить n-мерный тетраэдр, или симплекс, с $n + 1$ вершинами, попарно соединенными всеми возможными ребрами. Они, в свою очередь, образуют треугольники, которые собираются в тетраэдры и т. д. Теперь несложно определить циклы, границы и гомологию и опять же можно состряпать группу путем добавления (гомологических классов) циклов. Фактически мы получаем целую серию групп: одну для нульмерных циклов (точек), одну для одномерных циклов (отрезков), одну для двумерных циклов (треугольников) и т. д. до полной размерности пространства. Это нулевая, первая, вторая и т. д. гомологические группы пространства. Грубо говоря, они уточняют представление об отверстиях различных размерностей в пространстве: существуют ли они, сколько их и как они соотносятся друг с другом?

Это и есть гомология, и этого нам почти достаточно для понимания того, что *говорит* гипотеза Ходжа. Однако что нам на самом деле нужно, так это близкая к ней концепция *когомологии*. В 1893 г. Пуанкаре обратил внимание на любопытное совпадение в гомологии любого многообразия: список гомологических групп с начала и с конца читается одинаково. Для многообразия размерности 5, скажем, нулевая гомологическая группа совпадает с пятой, первая — с четвертой, а вторая — с третьей. Он понял, что это не может быть простым совпадением, и объяснил его двойственностью триангуляции, с которой мы уже встречались в главе 4 в связи с картами. Это второй вариант триангуляции, где каждый треугольник заменяется вершиной, каждая сторона, общая для двух треугольников, — ребром, соединяющим две вершины, а каждая точка — треугольником, как на рис. 9 в главе 4. Обратите внимание на то, что измерения появляются здесь в обратном порядке: двумерные треугольники превращаются в нульмерные точки, и наоборот; одномерные ребра остаются одномерными, потому что 1 находится в середине.

Оказывается, полезно различать два списка, хотя инварианты они выдают одни и те же. Когда все это обобщается и облекается в формальные термины, триангуляция исчезает, и дуальная триангуляция тоже теряет смысл. Остаются только две серии топологических инвариантов, именуемых гомологическими и когомологическими группами. Вообще, каждое понятие в гомологии имеет двойника, название которого обычно образуется от названия понятия путем добавления приставки «ко-». Таким образом, вместо циклов мы получаем коциклы, а вместо заявления о том, что два цикла гомологичны, говорим, что два коцикла когомологичны. Классы, о которых идет речь в гипотезе Ходжа, — это классы когомологий, которые представляют собой наборы когомологичных коциклов.

Гомология и когомология не сообщают нам всего, что мы хотели бы знать о форме топологического пространства, — различные пространства могут обладать идентичными гомологией и когомологией, — но дают немало полезной информации, а также обеспечивают системные рамки для его расчета и использования.

Алгебраическое многообразие — будь оно действительным или комплексным, проективным или нет — представляет собой топологическое пространство. Поэтому оно имеет форму. Чтобы выяснить об этой форме что-нибудь полезное, мы рассматриваем многообразие как топологи и вычисляем его гомологическую и когомологическую группы. Но естественными ингредиентами алгебраической геометрии являются не геометрические объекты вроде триангуляционных сеток и циклов, а вещи, которые проще всего описываются алгебраическими уравнениями. Вернитесь немного назад и взгляните еще раз на уравнение поверхности Куммера. Как это соотносится с триангуляцией? В формуле нет ничего, что указывало бы на треугольники.

Может быть, нам нужно начать сначала. Вместо треугольников нам следовало бы использовать естественный строительный материал для многообразий — подмногообразия, определенные дополнительными ограничивающими уравнениями. Теперь нам придется переопределять циклы: вместо набора треугольников с целыми ярлыками мы воспользуемся набором подмногообразий с такими ярлыками, которые лучше всего подойдут в данном случае. По различным причинам — по большей части потому, что, если использовать целые ярлыки, гипотеза Ходжа неверна, — разумным выбором будут рациональные числа. Вопрос Ходжа сводится к следующему: содержит ли новое определение гомологии и когомологии всю ту же информацию, что и топологическое определение? Если гипотеза верна, то алгебраический цикл — не менее острый инструмент топологии, чем когомологический резец. Если она неверна, то алгебраический цикл — всего лишь твердый тупой предмет.

Вот только… прошу прощения, я немного переборщил. Гипотеза утверждает, что достаточно воспользоваться определенным *типом* алгебраического цикла — того, что обитает в классе Ходжа. Чтобы объяснить это, нам потребуется еще один ингредиент в уже и без того густой смеси: анализ. Одной из важнейших концепций анализа является дифференциальное уравнение, которое представляет собой условие, наложенное на скорости изменения переменных (см. главу 8). Почти вся математическая физика XVIII, XIX и XX вв. моделирует реальность при помощи дифференциальных уравнений. По существу, это верно даже для XXI в. В 1930-е гг. эта идея привела Ходжа к целой группе новых методик. Сегодня все это называется теорией Ходжа. Она естественным образом связана с множеством других мощных методов в объединенной области анализа и топологии.

Идея Ходжа заключалась в том, чтобы использовать дифференциальное уравнение для распределения классов когомологий по типам. Каждый из них обладает дополнительной структурой, которую можно успешно применять при решении топологических задач. Определяются они при помощи дифференциального уравнения, появившегося впервые в конце XVIII в. в работе Пьера-Симона де Лапласа и известного, соответственно, как уравнение Лапласа. Основные работы Лапласа были посвящены небесной механике, движению и форме планет и их спутников, комет и звезд. В 1783 г. он работал над определением точной формы Земли. К тому времени уже было известно, что Земля — не сфера, что она сплющена у полюсов и представляет собой приплюснутый сфероид — как если сесть сверху на пляжный мяч. Но даже такое описание не отражает деталей. Лаплас нашел способ рассчитать форму Земли с любой заданной точностью на основании физической величины, представляющей гравитационное поле планеты: это не само поле, но его гравитационный потенциал. Это мера энергии, содержащейся в гравитационном поле, численная величина, определяемая в каждой точке пространства. Тяготение действует в том направлении, в котором потенциал уменьшается с максимальной скоростью, а абсолютное значение силы соответствует скорости уменьшения.

Гравитационный потенциал удовлетворяет уравнению Лапласа: грубо говоря, это означает, что в отсутствии вещества, т. е. в вакууме, среднее значение потенциала по очень маленькой сфере равно его значению в центре сферы. Это своего рода демократия: ваша ценность получается путем усреднения ценностей ваших соседей. Любое решение уравнения Лапласа называется гармонической функцией. Ходжа среди классов когомологий интересуют те, что имеют особые отношения с гармоническими функциями. Теория Ходжа и изучение этих типов помогли открыть глубокую и чудесную область математики: отношения между топологией пространства и специальным дифференциальным уравнением на этом пространстве.

Вот мы и у цели. Гипотеза Ходжа постулирует глубокую и мощную связь между тремя столпами современной математики: алгеброй, топологией и анализом. Возьмем любое многообразие. Чтобы разобраться в его форме (это топология с выходом на когомологические классы), выбираем частные случаи таких классов (анализ с выходом на классы Ходжа через дифференциальные уравнения). Эти частные случаи коголомологических классов могут быть реализованы с использованием подмногообразий (алгебра: добавьте несколько уравнений и внимательно посмотрите на алгебраические циклы). Иными словами, чтобы ответить на топологический вопрос («Какой формы эта штука?») для многообразия, следует перевести его в плоскость анализа, а затем решить средствами алгебры.

Почему это так важно? Гипотеза Ходжа — это предложение добавить в инструментарий специалиста по алгебраической геометрии два новых инструмента: топологические инварианты и уравнение Лапласа. В самом деле, если разобраться, то в этой гипотезе речь не идет о какой-то математической теореме: речь о новых инструментах. Если гипотеза верна, эти инструменты обретают новое значение и становятся потенциальным средством поиска ответов на бесчисленное количество вопросов. Разумеется, гипотеза может оказаться и ошибочной. Было бы обидно, но, если возможности наших инструментов ограничены, лучше знать об этом заранее, чем то и дело натыкаться на проблемы в самый неподходящий момент.

Теперь, когда мы оценили природу гипотезы Ходжа, можно посмотреть, какие у нас есть свидетельства в ее пользу. Что нам известно? Чрезвычайно мало.

В 1924 г., еще до того, как Ходж выдвинул свою гипотезу, Соломон Левшец доказал теорему, которая сводится к гипотезе Ходжа для второй (или двумерной) группы когомологий любого многообразия. При помощи рутинных методов алгебраической топологии можно показать, что из этого следует гипотеза Ходжа для размерностей 1, 2 и 3. Для многообразий более высоких размерностей известно лишь несколько частных случаев гипотезы Ходжа.

Первоначально Ходж сформулировал свою гипотезу в терминах целых маркеров (или индексов). В 1961 г. Майкл Атья и Фридрих Хирцебрух доказали, что для высших измерений эта версия гипотезы неверна. Поэтому сегодня мы формулируем гипотезу Ходжа с использованием рациональных коэффициентов: для этой версии у нас есть некоторое количество обнадеживающих данных. Самое сильное свидетельство в ее пользу состоит в том, что одно из наиболее глубоких ее следствий — еще более технически сложная теорема, известная как теорема об «алгебраичности локусов Ходжа», уже доказана *без опоры* на гипотезу Ходжа. Эдуардо Каттани, Пьер Делинь и Арольдо Каплан нашли соответствующее доказательство в 1995 г.

Наконец, в теории чисел имеется симпатичная гипотеза, аналогичная гипотезе Ходжа и получившая название гипотезы Тейта в честь Джона Тейта. Она связывает алгебраическую геометрию с теорией Галуа — совокупностью идей, доказывающих, что у полиномиальных уравнений пятой степени не существует явных решений, выражаемых формулой. Формулировка гипотезы Тейта достаточно сложна: в ней фигурирует еще один вариант когомологии. Есть причины надеяться, что гипотеза Тейта верна, хотя она не доказана. Но по крайней мере можно сказать, что у гипотезы Ходжа есть разумный родич, хотя как подступиться хоть к той, хоть к другой гипотезе, пока совершенно неясно.

Гипотеза Ходжа — одно из тех математических утверждений, которые почти нечем ни подтвердить, ни опровергнуть и у которых свидетельства и в ту и другую сторону не слишком убедительны. К тому же существует опасность, что гипотеза может оказаться попросту неверной. Возможно, существует многообразие с миллионом измерений, опровергающее гипотезу Ходжа по причинам, которые сводятся к серии неструктурированных расчетов, настолько сложных, что никто и никогда не сможет их провести. Если это так, то гипотеза Ходжа может оказаться ошибочной по совершенно глупой причине — просто так получилось, — но доказать это практически невозможно. Я знаю несколько специалистов по алгебраической геометрии, которые считают именно так. В этом случае обещанному миллиону долларов в обозримом будущем ничего не грозит.

16

Куда дальше?

Предсказывать очень трудно, особенно будущее. По легенде, так любили говорить знаменитый физик и нобелевский лауреат Нильс Бор и знаменитый бейсболист и спортивный менеджер Йоги Берра[44]. Правда, Берра, как утверждают, еще говорил так: «Имейте в виду, я никогда не говорил большей части того, что говорил».

Артур Кларк, знаменитый своими научно-фантастическими романами и фильмом «Космическая одиссея — 2001», был, помимо всего прочего, футурологом: он писал книги о будущем техники и общества. В его книге «Очертания будущего» (Profiles of the Future), написанной в 1962 г., среди прочих предсказаний можно найти следующие:

- к 1970 г. — расшифровка языка китов и дельфинов;
- к 1990 г. — создание термоядерного реактора;
- к 1990 г. — обнаружение гравитационных волн;
- к 2000 г. — колонизация планет.

Ничего подобного пока не произошло. Но, с другой стороны, у него были и удачные предсказания:

- к 1980 г. — приземление на другие планеты (хотя он, возможно, имел в виду высадку человека);
- к 1970 г. — машины-переводчики (слегка преждевременно, но сегодня машинный перевод существует в Интернете);
- к 1990 г. — индивидуальное радио (примерно эту роль сегодня исполняют мобильные телефоны).

Он также предсказывал, что к 2000 г. у нас будет глобальная библиотека, и сегодня это предсказание ближе к истине, чем можно было подумать еще несколько лет назад (это тоже одна из функций Интернета). С развитием облачных вычислений мы, возможно, когда-нибудь все станем пользователями одного и того же гигантского компьютера. При этом Кларк упустил из виду некоторые важнейшие тенденции, такие как расцвет компьютеров и генная инженерия, хотя ее-то он как раз предсказал, но на 2030 г. Учитывая спорные суммарные результаты предсказаний Кларка, со своей стороны я бы не рискнул предсказывать будущее великих математических задач сколько-нибудь подробно. Однако могу высказать кое-какие квалифицированные догадки, не сомневаясь, однако, что большинство из них окажутся в результате ошибочными.

Во введении я упоминал список Гильберта из 23 крупных проблем, озвученный в 1900 г. В большинстве своем они уже решены, и может показаться, что смелый призыв ученого «Мы должны знать, и мы будем знать» оправдался. Однако Гильберт сказал также: «В математике ни о чем нельзя утверждать, что мы никогда этого не узнаем». Курт Гедель расправился с этой идеей, доказав свою теорему о неполноте: некоторые математические задачи могут не иметь решения в рамках обычной логической математики. Их решение не просто невозможно, как извлечение квадратуры круга — они могут быть неразрешимы в том смысле, что для них не существует ни доказательства, ни опровержения. Вероятно, именно такая судьба ждет некоторые из сегодняшних великих задач математики. Но я был бы удивлен, если бы в их число вошла гипотеза Римана, и поражен, если бы кто-то смог доказать ее неразрешимость. С другой стороны, проблема P/NP-алгоритмов вполне может оказаться неразрешимой или подпадать под какое-то другое формальное ограничение вида «это не может быть сделано, потому что…». Есть в этой задаче что-то, знаете ли, *эдакое*…

Подозреваю, что к концу XXI в. у нас все-таки будут доказательства гипотезы Римана, гипотезы Берча–Свиннертон-Дайера и гипотезы массовой щели, а также опровержение гипотезы Ходжа и регулярности решений уравнения Навье–Стокса в трех измерениях. Мне кажется, что задача P/NP-алгоритмов пока останется нерешенной, но падет где-нибудь в XXII в. Хотя понятно, что кто-нибудь завтра же опровергнет гипотезу Римана, а через неделю докажет, что P не совпадает с NP.

Но, конечно, безопаснее ограничиваться общими наблюдениями, потому что мы всегда можем поучиться у истории. Так что я почти уверен: к тому моменту, когда семь задач тысячелетия наконец решат, многие из них будут восприниматься как мелкие исторические курьезы. «О, когда-то *это* считалось важным?» Так произошло с некоторыми проблемами из списка Гильберта. Кроме того, можно быть уверенным, что через 50 лет появится несколько крупных областей математики, которых сегодня не существует даже в проекте. Тогда выяснится, что кое-какие базовые примеры и некоторые рудиментарные теоремы в этих областях известны уже давно, но никто не догадывался, что эти случайные кусочки смальты представляют собой фрагменты прекрасной мозаичной картины — глубокой и значительной новой области математики. В свое время так произошло с теорией групп, матричной алгеброй, фракталами и теорией хаоса. Не сомневаюсь, что это произойдет еще не раз, потому что это стандартный путь развития математики.

Существует две основные движущие силы возникновения новых областей. С одной стороны, их порождает внутренняя структура самой математики, а с другой — они являются ответом на новые вопросы о внешнем мире; часто оба фактора действуют одновременно. Как у Пуанкаре процесс решения задачи включал в себя три этапа — подготовку, созревание и озарение, — так и отношения между математикой и ее приложениями не сводятся к простой схеме: физика ставит вопрос, математика дает на него ответ, и дело с концом. На самом деле мы видим сложную систему обмена вопросами и идеями: новые достижения в математике служат стимулами для дальнейших экспериментов, наблюдений или теорий, а те, в свою очередь, мотивируют новые математические исследования. И каждый узел этой сети оказывается, при ближайшем рассмотрении, самостоятельной сетью того же типа, но меньших масштабов.

Окружающий мир стал гораздо обширнее и богаче, чем прежде. До недавнего времени основным внешним источником вдохновения для математики была физика. Некоторые другие области науки тоже играли свою роль: биология и социология стимулировали развитие теории вероятностей и статистики, а философия заметно влияла на математическую логику. В будущем нам предстоит увидеть, как математика начнет все более активно взаимодействовать с биологией, медициной, компьютерными науками, финансами, экономикой, социологией и, очень возможно, политикой, а также киноиндустрией и спортом. Я подозреваю, что некоторые из ближайших к нам по времени великих задач возникнут в биологии, поскольку с ней уже установилась прочная связь. Одна из тенденций — взрывной рост возможностей по сбору биологических и биохимических данных. Так, сегодня небольшие геномы можно секвенировать при помощи устройства размером с флешку (методом нанопорового анализа). Очень скоро то же можно будет проделывать и с большими геномами при помощи тех же или других технологий. В любом случае большая часть соответствующих технологий уже существует.

Потенциально уже эти достижения могут в значительной мере изменить обстановку. Однако нужно еще разобраться, что все эти данные означают. Вообще-то, биология — наука не о данных, а о процессах. Эволюция — это процесс, как и деление клетки, развитие зародыша, зарождение раковой опухоли, поведение толпы, работа мозга и динамика глобальной экосистемы. Лучший известный на сегодня способ понять фундаментальные свойства процесса и разобраться в том, что в нем происходит, как и почему, — это математика. Так что скоро появятся великие задачи новых типов — как разворачивается динамика процесса в присутствии сложной, но очень конкретной организующей информации (ДНК-последовательностей); как генетические изменения взаимодействуют со средой, сдерживая эволюцию; как правила роста, деления, движения, адгезии и гибели клетки формируют развивающийся организм; как поток электронов и химических веществ в сети нервных клеток определяет восприятие и действия организма.

Вычислительные средства — еще один источник новой математики, успевший уже проявить себя. Обычно они воспринимаются как инструмент математических действий, но не стоит забывать, что математика в равной степени является инструментом понимания и организации вычислений. Этот двусторонний обмен приобретает все большее значение для благополучия и развития обеих областей — не исключено, что когда-нибудь в будущем они просто сольются воедино. Некоторые математики считают, что их с самого начала не следовало разделять. Из множества существующих уже сегодня в этой области тенденций на ум приходит вопрос о работе с очень большими массивами данных. Этот вопрос имеет отношение не только к ДНК, о чем уже упоминалось, но и к задаче предсказания землетрясений, к расчету эволюционных процессов, к проблемам глобального климата, фондового рынка, международных финансов и новых технологий. Наша задача — научиться использовать большие массивы данных для проверки и отладки математических моделей реального мира, которые в дальнейшем дадут нам в руки подлинный контроль над сложнейшими системами.

В отношении того, в чем лично я разбираюсь лучше всего, предсказания в основном негативные, но в то же время это подтверждает, что креативность математического сообщества по-прежнему не снижается. Все математики-исследователи время от времени чувствуют, что их предмет как будто обладает собственным сознанием. Задачи решаются так, как это нужно математике, а не математикам. Мы выбираем, какие вопросы рассматривать, но мы не можем выбирать, какие у этих вопросов должны быть ответы. Вообще, такое ощущение характерно для двух крупных школ, которые отличает разное отношение к природе математики. Последователи Платона считают, что «идеальные формы» математики ведут своего рода независимое существование «где-то там», в некоем собственном царстве, отличном от материального мира. (Существуют более тонкие формулировки, которые, вероятно, звучат более здраво, но суть именно в этом.) Вторые видят в математике общечеловеческую концепцию. Но, в отличие от большинства подобных концепций — законодательной системы, денег, этики, морали, математика представляет собой конструкцию с прочной логической основой. Существуют серьезные ограничения на то, какие утверждения вы можете или не можете предлагать остальным. Именно из-за этих ограничений возникает впечатление, что математика сама решает, что ей делать и как развиваться; они же создают у математиков ощущение, что их наука существует *сама по себе* вне зоны человеческой деятельности. Мне представляется, что платонизм — это описание не того, что есть математика на самом деле, а того, как *ощущает* математику человек, в ней работающий. Примерно так человек, увидевший розу, кровь или светофор, живо ощущает «красное». Философы называют подобные ощущения «первичными», некоторые из них даже считают, что наше ощущение свободы воли на самом деле представляет собой первичное ощущение того, как мозг принимает решения. Выбирая из нескольких вариантов, мы уверены, что действительно свободны в своем выборе, — хотя не исключено, что динамика мозга в каком-то смысле детерминирована. Тогда платонизм — это первичное ощущение участия в общечеловеческом процессе, ограниченном жесткими рамками логических построений.

Так что может показаться, что математика обладает собственным сознанием, даже если на самом деле она создается коллективной интеллектуальной деятельностью людей. История учит нас, что математическое сознание — в этом смысле — более изобретательно и удивительно, чем можно себе представить. Все это лишь подходы к моему главному утверждению: единственное, что можно с уверенностью предсказать в математике, это ее непредсказуемость. Важнейшие математические вопросы начавшегося столетия возникнут как естественные, иногда даже неизбежные, следствия накопления наших знаний о нынешних великих задачах математики. Однако почти наверняка это будут вопросы, которые никто сегодня не может даже вообразить. Это верно и правильно, и нам нужно этому радоваться.

17

Двенадцать задач на будущее

Не хочу оставить у вас неверное впечатление, что большинство математических задач (за исключением нескольких особенно сложных) уже решено. Математические исследования напоминают изучение новооткрытого материка. По мере того как расширяется уже исследованная область, становится длиннее и граница между известным и неизвестным. Я не утверждаю, что чем больше математических закономерностей мы открываем, тем меньше знаем. Я говорю, что чем больше математических закономерностей мы открываем, тем лучше представляем себе объемы непознанного. Но непознанное изменяется со временем: одни задачи уходят в прошлое, на горизонте появляются другие. А область известного только расширяется, если, конечно, не говорить о случайно утерянных документах.

Чтобы дать вам некоторое представление о том, чего мы *не знаем* в настоящий момент (помимо тех проблем, о которых мы уже говорили), я приведу 12 нерешенных задач, которые уже некоторое время ставят в тупик математиков всего мира. Я выбрал их таким образом, чтобы несложно было понять суть вопроса. Мы уже видели, что простота формулировок ничего не говорит о том, насколько легким или сложным может быть доказательство. Некоторые из этих проблем еще могут обернуться великими: это будет зависеть в основном не от ответа на вопрос, а от того, какие методы будут придуманы и применены для их решения и к чему соответствующие исследования в конце концов приведут.

Задача Брокара

Для любого целого числа n факториал $n!$ равен произведению

$$n \times (n-1) \times (n-2) \times \ldots \times 3 \times 2 \times 1.$$

Это число различных способов расставить по порядку n объектов. К примеру, английский алфавит, содержащий 26 букв, можно расставить

$$26! = 403\,291\,461\,126\,605\,635\,584\,000\,000$$

разными способами. В статьях, опубликованных в 1876 и 1885 гг., Анри Брокар отметил, что

$$4! + 1 = 24 + 1 = 25 = 5^2,$$
$$5! + 1 = 120 + 1 = 121 = 11^2,$$
$$7! + 1 = 5040 + 1 = 5041 = 71^2$$

представляют собой полные квадраты. Он не обнаружил других факториалов, которые при прибавлении единицы давали бы полный квадрат, и задался вопросом, существуют ли такие. Индийский гений-самоучка Шринивaса Рамануджан независимо задался этим же вопросом в 1913 г. В 2000 г. Брюс Берндт и Уильям Голуэй при помощи компьютера показали, что для факториалов чисел до 1 млрд других решений не существует.

Нечетные совершенные числа

Число является совершенным, если оно равно сумме всех его собственных делителей (т. е. чисел, на которые оно делится без остатка, включая единицу, но исключая само число). Примеры таких чисел:

$$6 = 1 + 2 + 3,$$
$$28 = 1 + 2 + 4 + 7 + 14.$$

Евклид доказал, что если число $2^n - 1$ простое, то число $2^{n-1}(2^n - 1)$ совершенно. Приведенные выше примеры соответствуют $n = 2, 3$. Простые числа такого вида называются простыми Мерсенна, их известно 47 штук, и самое большое из них $2^{43\,112\,609} - 1$ (кроме того, это самое большое известное простое число). Эйлер доказал, что все четные совершенные числа должны иметь такой вид, но никому еще не удалось отыскать хотя бы одно нечетное совершенное число или доказать, что их не существует. Померанс предложил нестрогое рассуждение, которое вроде бы указывает, что таких чисел действительно нет. Любое нечетное совершенное число должно удовлетворять нескольким жестким условиям. По величине оно должно быть не меньше 10^{300}, должно иметь простой делитель больше чем 10^8, его второй по величине простой делитель должен быть по крайней мере 10^4; кроме того, у него должно быть по крайней мере 75 простых делителей и по крайней мере 12 различных простых делителей.

Гипотеза Коллатца

Возьмем целое число. Если оно четное, разделим на 2. Если нечетное, домножим на 3 и прибавим 1. Повторим эту операцию бесконечное число раз. Что произойдет?

К примеру, можно начать с числа 12. Получим следующую последовательность:

$$12 \to 6 \to 3 \to 10 \to 5 \to 16 \to 8 \to 4 \to 2 \to 1,$$

после чего последовательность $4 \to 2 \to 1 \to 4 \to 2 \to 1$ будет повторяться бесконечно. Гипотеза Коллатца утверждает, что конечный результат будет одним и тем же, с какого бы числа мы ни начали. Гипотеза названа в честь Лотара Коллатца, предложившего ее в 1937 г., но имеет и множество других названий: гипотеза $3n + 1$, гипотеза градины, гипотеза Улама, проблема Какутани, гипотеза Туэйтса, алгоритм Хассе или сиракузская проблема.

Что делает эту задачу такой сложной, так это то, что нередко числа буквально взрываются. Так, если начать с 27, последовательность поднимется до 9232, но при этом все равно через 111 шагов сойдется к 1. Компьютерное моделирование подтверждает гипотезу для всех первоначальных чисел вплоть до $5{,}764 \times \times 10^{18}$. Доказано, что не существует циклов, за исключением $4 \to 2 \to 1$, в которых было бы меньше 35 400 шагов. Возможность того, что некоторое начальное число дает последовательность, содержащую все более крупные числа, разделенные более мелкими, не исключена. Илья Красиков и Джеффри Лагариас доказали, что для начальных величин вплоть до n по крайней мере $n^{0{,}84}$ из них со временем сходится к 1. Так что исключения, если они существуют, встречаются редко.

Существование правильного кубоида

Здесь в качестве начального пункта берется существование пифагоровых троек и формула для них, а затем вся проблема переводится в третье измерение. Эйлеров параллелепипед — это кубоид (блок в форме кирпича) с целыми ребрами, все грани которого имеют целые диагонали. Самый маленький параллелепипед Эйлера открыл в 1719 г. Пауль Хальке. Его ребра составляют 240, 117 и 4; диагонали граней равны 267, 244 и 125. Эйлер нашел формулы для таких прямоугольных параллелепипедов, аналогичные формуле для пифагоровых троек, но они выдают не все возможные решения.

Неизвестно, существует ли совершенный кубоид, т. е. существует ли такой параллелепипед Эйлера, главная диагональ которого тоже имеет целую длину. (Главная диагональ — это отрезок, соединяющий противоположные вершины прямоугольного параллелепипеда и проходящий сквозь его внутреннюю часть. Таких отрезка четыре, но все они равны по длине.) Известно, что формулы Эйлера не дают примера такого параллелепипеда. Он, если существует, должен удовлетворять нескольким условиям — к примеру, по крайней мере одно его ребро должно быть кратно 5, другое — 7, третье — 11, четвертое — 19. Компьютерные эксперименты показали, что длина одного из ребер должна быть не менее одного триллиона.

Есть достаточно близкие варианты. У прямоугольного параллелепипеда со сторонами 672, 153 и 104 главная диагональ целая, как и две из трех диагоналей граней. В 2004 г. Хорхе Сойер и Клиффорд Рейтер доказали, что существуют совершенные непрямоугольные параллелепипеды. Грани таких параллелепипедов представляют собой не прямоугольники, а параллелограммы, а сам параллелепипед как бы скошен на сторону. Ребра совершенного непрямоугольного параллелепипеда имеют длины 271, 106 и 103; малые диагонали граней равны 101, 266 и 255; большие диагонали граней — 183, 312 и 323; внутренние диагонали (а у такого параллелепипеда они все разные) имеют длины 374, 300, 278 и 272.

Гипотеза об одиночестве бегуна

Эта задача из трудной для понимания области математики, известной как теория диофантовых приближений. Сформулировал ее в 1967 г. Йорг Виллс. А название — гипотеза одинокого бегуна — придумал в 1998 г. Луис Годдин. Положим, что n бегунов бегают по кольцевой дорожке единичной длины с постоянной скоростью, причем скорости всех бегунов различны. Можно ли утверждать, что каждый из бегунов в какой-то момент времени окажется одиноким, т. е. будет находиться на расстоянии более $1/n$ от остальных? Разумеется, для разных бегунов это будут разные моменты времени. Гипотеза состоит в том, что ответ всегда «да»; на данный момент она доказана для $n = 4, 5, 6$ и 7.

Гипотеза Конвея о трекле

Трекл — это сеть, размещенная на плоскости таким образом, что каждые два ее ребра имеют ровно одну общую точку (см.

рис. 48). Встречаться они могут либо в вершине, либо в точке пересечения, но не то и другое одновременно. Если они пересекаются, то обязательно поперек; это значит, что ни одно из них не может целиком остаться по одну сторону от другого (а это могло бы произойти, если бы они, скажем, соприкасались). Джон Конвей в неопубликованной работе высказал гипотезу о том, что в любой сети такого рода число линий меньше или равно числу точек. В 2011 г. Радослав Фулек и Янош Пач доказали, что любая такая сеть с n точками имеет не более $1{,}428n$ линий.

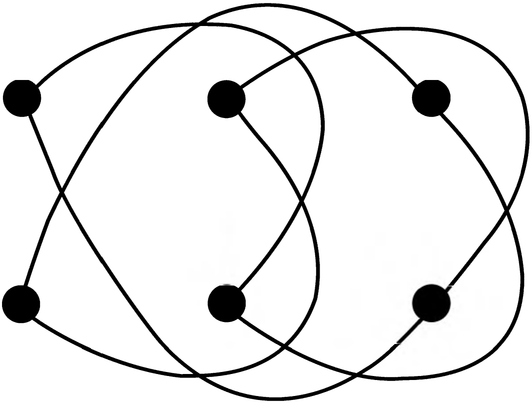


Рис. 48. Пример трекла

Иррациональность постоянной Эйлера

Не существует готовой «замкнутой» формулы для суммы гармонического ряда

$$H_n = 1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \frac{1}{5} + \ldots + \frac{1}{n}.$$

Более того, такой формулы, по всей вероятности, не существует. Однако существует прекрасная ее аппроксимация: по мере того как n увеличивается, H_n стремится к $\log n + \gamma$.

Здесь γ — постоянная Эйлера, численно равная примерно 0,5772156649. Эйлер вывел эту формулу в 1734 г., а Лоренцо Маскерони изучал постоянную в 1790 г. Ни тот, ни другой не использовали символ γ.

Постоянная Эйлера — одно из тех странных чисел, которые время от времени возникают в математике (вспомните π и e); у них нет красивого или простого выражения, они то и дело появляются в самых разных местах, но при этом складывается впечатление, что они существуют сами по себе. В главе 3 мы убедились, что и π, и e трансцендентны: они не являются решениями каких-либо алгебраических уравнений с целыми коэффициентами. Они иррациональны: не выражаются точными дробями. Многие математики считают, что постоянная Эйлера трансцендентна, но мы даже не знаем наверняка, иррациональна ли она. Если все же $\gamma = p/q$ для целых p и q, то q равняется по крайней мере $10^{242\,080}$.

Постоянная Эйлера важна во многих областях математики — от римановой дзета-функции до квантовой теории поля. Она появляется во многих ситуациях и в многочисленных формулах. Поэтому просто возмутительно, что мы не можем решить, рациональна ли она!

Действительные квадратичные числовые поля

В главе 7 мы видели, что одни алгебраические числовые поля имеют единственное разложение на простые множители, а другие — нет. Лучше всего изучены квадратичные алгебраические числовые поля, полученные путем извлечения квадратного корня из некоего числа d, которое не является полным квадратом, более того, не имеет делителей — полных квадратов. Соответствующее кольцо алгебраических целых чисел, состоящее из всех чисел вида $a+b\quad d$, где a и b — целые числа, если d не имеет вид $4k + 1$, и либо целые, либо нечетные целые, деленные на 2, если d имеет такой вид.

Если d отрицательно, то мы знаем, что разложение на простые множители является единственным ровно для девяти чисел:

–1, –2, –3, –7, –11, –19, –43, –67 и –163. Доказательство единственности в этих случаях относительно понятно, но вот поиск других таких чисел очень сложен. В 1934 г. Ганс Хайльбронн и Эдвард Линфут показали, что к этому списку можно добавить не более одного отрицательного целого числа. Курт Хегнер в 1952 г. предложил доказательство полноты списка, но считалось, что в этом доказательстве есть пробел. В 1967 г. Гарольд Старк нашел полное доказательство, заметив при этом, что оно незначительно отличается от доказательства Хегнера, т. е. что пробел не имел значения. Примерно в то же время Алан Бейкер нашел еще одно доказательство.

Случай, когда d положительно, совсем не такой. Разложение на простые множители единственно для гораздо большего числа значений d. Только до 50 это 2, 3, 5, 6, 7, 11, 13, 14, 17, 19, 21, 22, 23, 29, 31, 33, 37, 38, 41, 43, 46, 47. Компьютерные расчеты позволяют получить еще много значений. Насколько нам известно, может существовать бесконечно много положительных значений d, соответствующее которым квадратичное числовое поле однозначно раскладывается на простые множители. Эвристический анализ, проведенный Коэном и Ленстрой, позволяет предположить, что примерно три четверти всех положительных d, по идее, должны определять числовые поля с однозначными разложениями. Проблема в том, чтобы доказать, что эти наблюдения верны.

Муравей Лэнгтона

Годы идут, и становится все более очевидным, что традиционные методы математического моделирования уже не справляются с задачами, которые ставит перед собой человечество: моделированием глобальной финансовой системы, динамики экосистем, роли генов в росте и развитии живых организмов. Во многие из этих систем входит гигантское количество действующих «лиц» — людей, компаний, организмов, генов, взаимодействующих между собой. Нередко эти взаимодействия можно смоделировать при помощи достаточно простых правил. В последние 30 лет получил развитие новый тип модели, который пытается разобраться с поведением подобных систем, что называется, «в лоб». К примеру, чтобы понять, как 100 000 человек будут вести себя на стадионе, мы не станем усреднять их и превращать в своего рода человеческую жидкость, течение которой затем следует рассматривать. Нет, мы строим компьютерную модель из 100 000 отдельных модулей, накладываем на них подходящие ограничения, устанавливаем правила и запускаем процесс моделирования, чтобы посмотреть, что будет делать эта компьютерная толпа. Такого рода модели в математике называют сложными системами.

Чтобы дать вам некоторое представление об этой новой и очень интересной области математики, я опишу одну из простейших сложных систем и объясню, почему мы не понимаем ее до конца. Эта система называется муравьем Лэнгтона. Кристофер Лэнгтон был одним из первых сотрудников Института Санта-Фе, который основали в 1984 г. физики Джордж Коуэн, Марри Гелл-Ман и другие для развития теории и приложений сложных систем. Лэнгтон придумал своего муравья в 1986 г. Технически это клеточный автомат, система клеток квадратной решетки, состояния которых обозначаются цветом. На каждом временнóм шаге цвет каждой клетки изменяется в соответствии с цветом соседних с ней клеток.

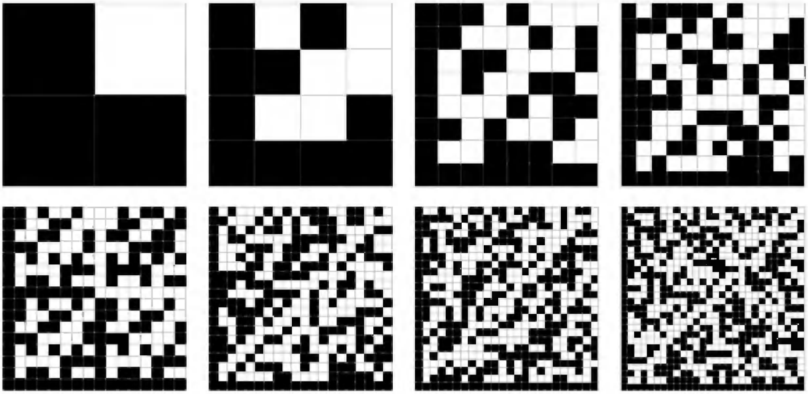
Правила просты до нелепости. Муравей живет на бесконечной квадратной решетке из клеток, и первоначально все они белые. Он всегда носит с собой неиссякаемый горшочек с черной краской и такой же горшочек с белой краской. Он может идти на север, на восток, на юг или на запад. Из соображений симметрии скажем, что первый шаг он делает на север. В каждый момент времени муравей смотрит на цвет клетки, в которой оказался, и перекрашивает ее из черной в белую или из белой в черную. Если клетка была белой, то после перекрашивания муравей поворачивает на 90° направо и делает один шаг вперед. Если клетка была черной, то он поворачивает на 90° налево и делает то же самое. И так до бесконечности.

Если вы смоделируете поведение муравья, то сначала он будет рисовать простой симметричный узор из белых и черных квадратов. Время от времени он возвращается на клетку, где уже был, но петля при этом не замыкается, потому что цвет клетки изменился, и муравей повернет в другую сторону. Моделирование продолжается, и рисунок становится хаотичным и случайным. При этом в нем невозможно различить никаких закономерностей: в основе своей это просто беспорядок. На этой стадии можно подумать (и вполне здраво), что такое хаотичное поведение будет продолжаться бесконечно. В конце концов, вернувшись в хаотично раскрашенный регион, муравей непременно сделает серию хаотичных шагов. Если вы будете продолжать моделирование, то следующие примерно 10 000 шагов подтвердят ваше предположение. Однако затем, если вы будете настойчивы, проявится закономерность. В движениях муравья возникнет повторяющийся цикл из 104 шагов, в результате которого он проходит две клетки по диагонали. После этого он будет двигаться, прорисовывая широкую диагональную полосу из черных и белых клеток, которую иногда называют магистралью, и так до бесконечности (см. рис. 49).

Рис. 49. Магистраль муравья Лэнгтона

Все описанное до сих пор может быть доказано по всей строгости просто последовательным перебором муравьиных шагов. Это будет достаточно длинное доказательство — список из 10 000 шагов, — но все же доказательство. Но математика системы станет более интересной, если мы зададимся чуть более общим вопросом. Что если еще до начала движения муравья мы перекрасим некоторое конечное число клеток решетки в черный цвет? Мы можем выбрать для этого любые клетки: это может быть случайный набор, черный квадрат или Мона Лиза. Их может быть миллион, или миллиард, или еще больше, но не бесконечное количество. Что произойдет?

Обычное движение муравья резко меняется при встрече с любой из новых черных клеток. Он может долго бродить окрест, рисуя сложные орнаменты и раз за разом перерисовывая их заново… Но во всех до сих пор предпринятых попытках, какой бы ни была первоначальная конфигурация, в конце концов муравей непременно принимался за строительство магистрали при помощи все того же 104-шагового цикла. Всегда ли это происходит? Является ли магистраль единственным «аттрактором» движения муравья? Никто не знает. Это одна из фундаментальных нерешенных задач теории сложности. Максимум,



Источник: http://mathworld.wolfram.com/HadamardMatrix.html

Рис. 50. Матрица Адамара размеров 2, 4, 8, 12, 16, 20, 24 и 28 что нам известно, — это то, что, какой бы ни была первоначальная конфигурация черных клеток, муравей не останется навечно в пределах ограниченной области поля.

Гипотеза Адамара

Матрица Адамара, названная в честь Жака Адамара, представляет собой квадратную матрицу из нулей и единиц, такую, что в любых двух ее рядах или столбцах половина элементов совпадает, а другая половина — отличается. На рис. 50 можно увидеть матрицы размеров 2, 4, 8, 12, 16, 20, 24 и 28, где 0 и 1 обозначены черным и белым цветом. Такие матрицы появляются во многих математических задачах и в компьютерных науках, в первую очередь в теории кодирования. (В некоторых приложениях, в том числе в задаче, которой первоначально занимался Адамар, белые квадраты соответствуют –1, а не 0.)

Адамар доказал, что подобные матрицы могут существовать только при $n = 2$ или n, кратном 4. Теорема Пейли 1933 г. доказывает, что матрица Адамара существует всегда для n, кратного 4 и равного $2^a(p^b + 1)$, где p — нечетное простое число. Из чисел, кратных 4, под эту теорему не подпадают 92, 116, 156, 172, 184, 188, 232, 236, 260, 268 и другие, более крупные значения n. Гипотеза утверждает, что матрица Адамара существует любых размеров, кратных 4. В 1985 г. К. Савад нашел матрицу размера 268. Есть и другие числа, не удовлетворяющие условию теоремы Пейли, с которыми уже разобрались. В 2004 г. Хади Харагани и Бехруз Тайфех-Резайе нашли матрицу Адамара размера 428, и теперь минимальное значение n, для которого она неизвестна, составляет 668.

Уравнение Ферма–Каталана

Это диофантово уравнение $x^a + y^b = z^c$, где a, b и c — положительные целые числа, показатели степени. Я назову это уравнение уравнением Ферма–Каталана, потому что его решения имеют отношение как к Великой теореме Ферма (см. главу 7), так и к гипотезе Каталана (см. главу 6). Если a, b и c малы, ненулевые целые решения не особенно удивительны. К примеру, если все они равны 2, мы имеем уравнение Пифагора, которое, как известно со времен Евклида, имеет бесконечно много решений. Так что основной интерес представляют те случаи, когда показатели степени велики. Формально они являются «большими», когда $s = 1/a + 1/b + 1/c$ меньше 1. Известно лишь десять больших решений уравнения Ферма–Каталана:

$$
\begin{aligned}
&1 + 2^3 = 3^2,\\
&2^5 + 7^2 = 3^4,\\
&7^3 + 13^2 = 2^9,\\
&2^7 + 17^3 = 71^2,\\
&3^5 + 11^4 = 122^2,\\
&17^7 + 76\,271^3 = 21063928^2,\\
&1414^3 + 2213459^2 = 65^7,\\
&9262^3 + 15312283^2 = 113^7,\\
&43^8 + 96222^3 = 30042907^2,\\
&33^8 + 1549034^2 = 15613^3.
\end{aligned}
$$

Первое из этих решений считается большим, потому что $1 = 1^a$ для любого a и для $a = 7$ в том числе. Гипотеза Ферма–Каталана утверждает, что для больших s уравнение Ферма–Каталана имеет лишь конечное число целых взаимно простых решений. Основной результат доказали в 1997 г. Анри Дармон и Лоик Мерель: не существует решений, в которых $c = 3$, а a и b равны и больше 3. Больше почти ничего не известно. Дальнейший прогресс, судя по всему, зависит от поразительной новой гипотезы, речь о которой пойдет ниже.

Гипотеза ABC

В 1983 г. Ричард Мейсон обратил внимание на то, что один случай Великой теоремы Ферма никем никогда не рассматривался: речь идет о первых степенях. Иными словами, об уравнении $a + b = c$.

На первый взгляд, эта мысль абсолютно бессмысленна. Чтобы решить это уравнение для любой из трех переменных, выразив ее через две остальные, не нужны большие познания в алгебре. К примеру, $a = c - b$. Однако есть еще контекст, который меняет все. Мейсон понял, что, если задать в отношении a, b и c правильные вопросы, все становится намного глубже. Результатом этой необыкновенной идеи стала новая гипотеза теории чисел с далекоидущими последствиями. Будучи доказанной, она помогла бы математикам разобраться с множеством нерешенных на данный момент задач и найти более качественные и простые доказательства некоторых крупнейших теорем теории чисел. Речь идет о гипотезе ABC, в пользу которой говорит огромное количество численных свидетельств. Основана она на свободной аналогии между целыми числами и многочленами.

Евклид и Диофант знали рецепт для пифагоровых троек, который мы сегодня записываем в виде формулы (см. главу 6). Можно ли применить ту же уловку к другим уравнениям? В 1851 г. Жозеф Лиувилль доказал, что для уравнения Ферма для степеней 3 и выше подобной формулы не существует. Мейсон применил аналогичные рассуждения к более простому уравнению

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

для трех многочленов. Вроде бы это чрезмерно, ведь все решения можно найти при помощи элементарной алгебры. Тем не менее главный результат элегантен и далеко не очевиден: если каждый многочлен имеет делитель, который представляет собой полный квадрат, куб или более высокую степень, то уравнение не имеет решений.

Теоремы о многочленах часто имеют аналоги среди теорем о целых числах. В частности, неприводимые многочлены соответствуют простым числам. Естественный аналог теоремы Мейсона о многочленах в области целых чисел выглядит так. Пусть

$a + b = c$, где a, b и c — целые числа без общих делителей. Тогда простых делителей у каждого из чисел a, b и c меньше, чем *различных* простых делителей у числа abc. К несчастью, простые примеры показывают, что это неправда. В 1985 г. Дэвид Массер и Джозеф Эстерле модифицировали это утверждение и предложили вариант гипотезы, который не противоречит никаким известным примерам. Очень возможно, что их гипотеза ABC на данный момент является крупнейшей открытой проблемой в математике[45]. Если бы завтра кто-то доказал гипотезу ABC, многие глубокие и сложные теоремы, доказанные в последние десятилетия с громадными усилиями, получили бы новые простые доказательства. Еще одним следствием стала бы гипотеза Маршалла Холла: разность между любым полным кубом и любым полным квадратом должна быть достаточно большой. Наконец, еще одно потенциальное приложение гипотезы ABC — задача Брокара, первая в этой главе. В 1993 г. Мариус Оверхольт доказал, что если гипотеза ABC верна, то уравнение Брокара имеет конечное число решений.

Одно из самых интересных следствий гипотезы ABC связано с гипотезой Морделла. Фальтингс доказал это достаточно хитрым способом, но его результат был бы более убедительным, если бы нам известна была еще одна вещь: предел размера решений. Тогда существовал бы алгоритм поиска их всех. В 1991 г. Ноам Элкис показал, что частный случай гипотезы ABC, в которой различные постоянные ограничены, подразумевает такое улучшение теоремы Фальтингса. Лоран Море-Бэйи показал, что обратное верно. Из достаточно серьезных ограничений на величину решений *всего одного* диофантова уравнения, $y^2 = x^5 - x$, следует полная гипотеза ABC. Вообще, гипотеза ABC, хотя и не так хорошо известна, как многие другие нерешенные задачи, безусловно, является одной из великих математических задач. Как сказали Гранвиль и Томас Такер, ее решение оказало бы «необычайное влияние на наши представления о теории чисел. Было бы здорово доказать или опровергнуть ее»[46].

Глоссарий

Алгебраическое целое число. Комплексное число, удовлетворяющее полиномиальному уравнению с целыми коэффициентами и коэффициентом при наибольшей степени 1. К примеру, $i\ \ 2$, удовлетворяющее уравнению $x^2 + 2 = 0$.

Алгебраическое число. Комплексное число, удовлетворяющее полиномиальному уравнению с целыми коэффициентами или эквивалентными рациональными коэффициентами. К примеру, $i\ \ 2/3$ удовлетворяет уравнению или эквивалентно $9x^2 + 2 = 0$.

Алгебраическое многообразие. Множество в многомерном пространстве, определяемое системой алгебраических уравнений.

Алгоритм. Определенная процедура решения задачи, гарантированно приводящая к ответу.

Арифметическая прогрессия. Последовательность чисел, в которой каждое следующее число равно предыдущему плюс некая постоянная величина, разность прогрессии. Пример такой последовательности: 2, 5, 8, 11, 14… с разностью 3.

Асимптотический. Две величины, определенные через одну переменную, асимптотически равны, если по мере произвольного роста переменной их отношение все сильнее приближается к 1.

Бозон Хиггса. Элементарная частица, существование которой объясняет, почему все частицы обладают массой. О его открытии на Большом адронном коллайдере было объявлено в июле 2012 г.

Вектор. В механике величина, которая характеризуется как размером, так и направлением.

Верхняя граница. Конкретное число, гарантированно большее, чем некая искомая величина.

Вихрь. Жидкость, кружащаяся в водовороте. Может быть любого размера, в том числе очень маленького.

Волна. Возмущение, которое движется сквозь среду — твердое тело, жидкость или газ, не оставляя после себя в среде никаких постоянных изменений.

Вращение (поворот). На плоскости: преобразование, при котором все точки сдвигаются на один и тот же угол вокруг фиксированного центра. В пространстве: преобразование, при котором все точки сдвигаются на один и тот же угол вокруг фиксированной прямой — оси вращения.

Время разрушения решения. Время, после которого решение дифференциального уравнения прекращает существовать.

Гомология (группа). Топологический инвариант пространства, определенный замкнутыми петлями. Две петли гомологичны, если их разность представляет собой границу топологического диска.

Гомотопия (группа). Топологический инвариант пространства, определенный замкнутыми петлями. Две петли гомотопичны, если любая из них может быть непрерывно преобразована во вторую.

Гранецентрированная кубическая решетка. Повторяющаяся в пространстве совокупность точек. Кубики ставятся рядами и один на другой, образуя как бы трехмерную шахматную доску, а затем берутся их вершины и центры всех шести граней (см. рис. 17, 19).

Граница. Край определенной области.

Группа. Абстрактная алгебраическая структура, включающая в себя множество и правило комбинирования двух любых элементов множества, соответствующее трем условиям: в нем выполняется сочетательный закон, существует единичный элемент и каждому элементу соответствует обратный элемент.

Действительное число. Любое число, которое может быть выражено десятичной дробью, возможно, бесконечной. Пример: $\pi = 3{,}14159265358979 32385\ldots$

Дзета-функция. Комплексная функция, введенная Риманом и представляющая простые числа аналитически. Определяется рядом

$$\zeta(s)=\frac{1}{1^s}+\frac{1}{2^s}+\frac{1}{3^s}+\frac{1}{4^s}+\frac{1}{5^s}+\frac{1}{6^s}+\frac{1}{7^s}+\ldots,$$

который сходится, если действительная часть s больше единицы. Это определение может быть расширено на все комплексные s при помощи процесса, известного как аналитическое продолжение.

Динамическая система. Любая система, которая изменяется во времени по определенным правилам. К примеру, движение планет в Солнечной системе.

Диофантово уравнение. Уравнение, решения которого должны быть рациональными числами.

Диск (топологический). Область на поверхности, которую можно непрерывно преобразовывать в окружность вместе с тем, что у нее внутри.

Дифференциальное уравнение. Уравнение, в котором функция соотносится со скоростью ее изменения.

Дифференциальное уравнение в частных производных. Дифференциальное уравнение, в котором фигурируют скорости изменения некой функции по отношению к двум или более различным переменным (часто это пространство и время).

Додекаэдр. Многогранник, гранями которого являются 12 правильных пятиугольников (см. рис. 38).

Двойственная сеть. Сеть, полученная из данной сети. Чтобы получить ее, каждую область первоначальной сети следует заменить точкой и соединить эти точки ребрами, если соответствующие области граничат (см. рис. 10).

Единственность разложения на простые множители. Свойство, согласно которому любое число может быть записано как произведение простых множителей единственным способом с точностью до порядка записи множителей. Это верно для целых чисел, но не всегда верно в более общих алгебраических системах.

Идеальное число. Число, которое не входит в данную систему алгебраических чисел, но связано с этой системой так, что восстанавливает единственность разложения на простые множители в случаях, когда это свойство нарушается. В современной алгебре заменен идеалом — особым подмножеством той же системы.

Импульс. Произведение массы на скорость.

Индукция. Общий метод доказательства теорем о натуральных числах. Если какое-то свойство истинно для 0 и из его истинности для любого натурального n следует его истинность для $n + 1$, это свойство истинно для всех натуральных чисел.

Интеграл. Операция исчисления, при которой, по существу, складывается очень большое количество очень маленьких составляющих. Интеграл функции равен площади под ее графиком.

Иррациональное число. Действительное число, которое не является рациональным, т. е. не может быть записано в виде p/q, где p и q — целые числа и $q \neq 0$. Примерами могут служить 2 и π.

Калибровочная симметрия. Группа местных симметрий системы уравнений: преобразования переменных в разных точках пространства может быть различными, но, если обеспечить уравнениям компенсирующее изменение с разумным физическим обоснованием, любое решение системы остается решением.

Калибровочная теория. Квантовая теория поля с группой калибровочных симметрий.

Квадрат. Результат умножения числа на самое себя. К примеру, квадрат 7 равен $7 \times 7 = 49$, обозначается 7^2.

Квадратное уравнение. Любое уравнение $ax^2 + bx + c = 0$, где x — неизвестное, а a, b, c — константы.

Квантовая теория поля. Квантовомеханическая теория величины, которая пронизывает пространство и может иметь (и обычно имеет) разные значения в разных его местах.

Квантово-волновая функция. Математическая функция, определяющая свойства квантовой системы.

Класс E. Алгоритм, время работы которого для входа размера n пропорционально n-й степени некоей постоянной величины.

Класс P. Алгоритм, время работы которого пропорционально некоей постоянной степени размера входа.

Класс не-P. Не класс P.

Класс NP. Задача, для которой предлагаемое решение может быть проверено (но необязательно найдено) при помощи алгоритма класса P.

Класс Ходжа. Когомологический класс циклов на алгебраическом многообразии с особыми аналитическими свойствами.

Когомологическая группа. Абстрактная алгебраическая структура, связанная с топологическим пространством, аналогичная гомологической группе, но «двойственная» ей.

Комплексный анализ. Анализ — логически строгие вычисления, осуществляемые при помощи комплексных функций комплексного переменного.

Комплексное число. Число вида $a + bi$, где i — корень квадратный из –1, а a и b — действительные числа.

Конгруэнтное число. Число, которое может быть общей разностью последовательности трех квадратов рациональных чисел.

Контрпример. Пример, опровергающий некое утверждение. Так, 9 может служить контрпримером к утверждению «все нечетные числа простые».

Корень из единицы. Комплексное число ζ, некоторая степень которого $\zeta^k = 1$ (см. рис. 7 и прим. 33).

Коэффициент. В многочлене, таком как $6x^3 - 5x^2 + 4x - 7$, коэффициентами являются числа 6, –5, 4, –7, на которые домножаются различные степени x.

Координата. Одно из чисел в списке, определяющем положение точки на плоскости или в пространстве.

Косинус. Тригонометрическая функция угла, определяемая как $\cos A = a/c$ на рис. 51.

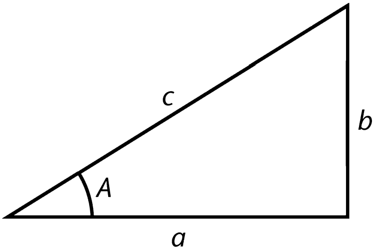


Рис. 51. Косинус (*a*/*c*), синус (*b*/*c*) и тангенс (*b*/*a*) угла *A*

Кривизна. Мера искривления пространства в окрестности данной точки. Сфера обладает положительной кривизной, плоскость — нулевой кривизной, а седловидная поверхность — отрицательной.

Круговое число, круговое целое число. Сумма степеней комплексного корня из единицы с рациональными либо целыми коэффициентами.

Куб. Число, умноженное на себя и еще раз на себя. К примеру, куб 7 равен $7 \times 7 \times 7 = 343$. Обычно записывается как 7^3.

Кубическое уравнение. Любое уравнение вида $ax^3 + bx^2 + cx + d = 0$, где x — неизвестное, а a, b, c, d — постоянные.

L-функция Дирихле. Обобщение дзета-функции Римана.

Логарифм. Натуральный логарифм x (обозначается $\log x$) — это степень, в которую нужно возвести e (= 2,71828...), чтобы получить x. Иными словами, $e^{\log x} = x$.

Логарифмический интеграл. Функция $Li(x) = \int_0^x \frac{dt}{\log t}$.

Максимум. Наибольшее значение чего-либо.

Минимальный контрпример. Математический объект, не обладающий неким желаемым свойством, причем в определенном смысле минимально возможный такой объект. К примеру, карта, которую невозможно раскрасить в четыре краски и состоящая притом из минимального числа областей, при котором это невозможно. Минимальные контрпримеры часто бывают гипотетическими, а цель математика при этом — доказать, что их не существует.

Минимум. Наименьшее значение чего-либо.

Многогранник. Тело, граница которого состоит из конечного числа многоугольников.

Многообразие. Многомерный аналог гладкой поверхности; форма в пространстве, определенная системой полиномиальных уравнений.

Многоугольник. Плоская фигура, граница которой состоит из конечного числа отрезков прямых.

Многочлен. Алгебраическое выражение, к примеру $6x^3–5x^2 + + 4x – 7$, в котором различные степени переменной x умножаются на константы и складываются.

Множество. Набор (математических) объектов. К примеру, множество целых чисел.

Модульная арифметика. Арифметическая система, в которой числа, кратные некоему заданному числу, называемому *модулем,* рассматриваются как равные нулю.

Момент импульса. Мера вращения тела.

Натуральное число. Любое из чисел 1, 2, 3…

Неевклидова геометрия. Альтернатива геометрии Евклида, в которой все обычные свойства точек и прямых сохраняются, за исключением допущения о существовании единственной прямой, параллельной данной и проходящей через заданную точку. Существует две разновидности неевклидовой геометрии: эллиптическая и гиперболическая.

Непрерывное преобразование. Преобразование пространства, при котором точки, расположенные очень близко друг к другу, не растаскиваются на большое расстояние.

Неприводимый многочлен. Многочлен, который нельзя получить при перемножении двух многочленов меньших степеней.

Неустойчивое состояние. Состояние динамической системы, к которому она не может вернуться после небольшого возмущения.

Неустранимая конфигурация. Элемент списка подсетей, по крайней мере одна из которых должна обязательно присутствовать в любой сети на плоскости.

NP-полная задача. Конкретная задача NP-класса, такая что если для ее решения существует алгоритм класса P, то *любая* задача класса NP может быть решена при помощи алгоритма класса P.

Нуль (функции). Если f — функция, то x является нулем f, если $f(x) = 0$.

Общая теория относительности. Теория гравитации Эйнштейна, в которой сила тяготения рассматривается как кривизна пространства-времени.

Оператор. Особый вид функции A, который при приложении к вектору v дает другой вектор Av. Должен удовлетворять условиям линейности: $\mathrm{A}(v + w) = \mathrm{A}v + \mathrm{A}w$ и $\mathrm{A}(av) = a\mathrm{A}(v)$ для любой постоянной a.

Оптимизация. Нахождение максимума или минимума некой функции.

Ось вращения. Фиксированная прямая, вокруг которой вращаются объекты.

Отношение. Отношение двух чисел a и b есть a/b.

Параллельный перенос. Преобразование пространства, при котором все точки сдвигаются в одном и том же направлении на одно и то же расстояние.

Переменная. Величина, которая может принимать любое значение в определенных пределах.

Периодичность. Бесконечная повторяемость одного и того же поведения.

Петля. Замкнутая кривая в топологическом пространстве.

Пифагорова тройка. Три натуральных числа a, b, c, такие что $a^2 + b^2 = c^2$. К примеру, $a = 3$, $b = 4$, $c = 5$. По теореме Пифагора такие числа образуют стороны прямоугольного треугольника.

Плоский тор. Тор, полученный отождествлением противоположных сторон квадрата, естественная геометрия которого имеет нулевую кривизну (см. рис. 12).

Поверхность. Форма в пространстве, полученная путем объединения областей, топологически эквивалентных внутренней части круга. Примеры: сфера и тор.

Показатель степени. Число, показывающее, в какую степень возводится переменная x. Для x^7 показатель степени — 7.

Поле скоростей. Функция, определяющая вектор скорости в каждой точке пространства. К примеру, в потоке жидкости вектор скорости может быть определен в каждой точке, и, как правило, в разных точках он разный.

Порядок кривой. Число раз, которые кривая оборачивается против часовой стрелки вокруг выбранной точки.

Последовательность. Список чисел в определенном порядке. К примеру, последовательность 1, 2, 4, 8, 16… степеней двойки.

Постоянная Эйлера. Специальное число, обозначаемое γ и приблизительно равное 0,57721 (см. прим. 41).

Построение при помощи линейки и циркуля. Любое геометрическое построение, которое можно реализовать при помощи только неразмеченной линейки и циркуля-измерителя (строго говоря, двух измерителей).

Поток Риччи. Уравнение, описывающее изменение кривизны пространства во времени.

Правильный многогранник. Многогранник, граница которого состоит из одинаковых правильных многоугольников, одинаково организованных возле каждой вершины. Евклид доказал, что существует ровно пять правильных многогранников.

Правильный многоугольник. Многоугольник, у которого все стороны имеют одинаковую длину, а все углы равны (см. рис. 4).

Преобразование. Еще одно слово, обозначающее «функцию»; используется обычно в тех случаях, когда задействованные переменные представляют собой точки в некотором пространстве. К примеру, «повернуть вокруг центра на 90°» — это преобразование квадрата.

Проективная геометрия. Разновидность геометрии, в которой параллельных прямых не существует: любые две прямые пересекаются в точке. Получается из евклидовой геометрии путем добавления новой «прямой в бесконечности».

Простое число. Натуральное число, большее 1, которое невозможно получить перемножением двух меньших натуральных чисел. Первыми простыми числами являются 2, 3, 5, 7, 11, 13.

Простой идеал. Аналог простого числа для алгебраических числовых систем.

Пятиугольник. Многоугольник с пятью сторонами.

Разложение на простые множители. Процесс, при котором число записывается в виде произведения его простых делителей. К примеру, разложение числа 60 выглядит как $2^2 \times 3 \times 5$.

Размерность. Число координат, необходимых для определения положения точки в данном пространстве. К примеру, размерность плоскости равна 2, а размерность пространства, в котором мы живем (по крайней мере с точки зрения геометрии Евклида) равна 3.

Ранг. Наибольшее число независимых рациональных решений уравнения, определяющего эллиптическую кривую. «Независимых» означает, что они не могут быть получены из других решений при помощи стандартного геометрического построения, которое из комбинации любых двух решений дает третье (см. рис. 25).

Рациональное число. Действительное число вида p/q, где p и q — целые числа и $q \neq 0$. Пример: $22/7$.

Решетка. На плоскости: множество точек, расположение которых повторяется в двух независимых направлениях, как узор на обоях (см. рис. 26). В пространстве: множество точек, расположение которых повторяется в трех независимых направлениях, как атомы в кристалле.

Решетчатая укладка. Набор одинаковых кружков или шариков, центры которых образуют решетку.

Род. Число отверстий в поверхности.

Ромбический додекаэдр. Многогранник, граница которого состоит из 12 одинаковых ромбов — параллелограммов с одинаковыми сторонами (см. рис. 15).

Ряд. Выражение, в котором складывается много — часто бесконечно много — величин.

Сводимая (сократимая) конфигурация. Часть сети, для которой характерно следующее: если сеть, полученную при ее удалении, можно раскрасить в четыре краски, то это можно сделать и с первоначальной сетью.

Симметрия. Преобразование некоторого объекта, при котором его форма в целом не меняется. К примеру, поворот квадрата на 90°.

Сингулярность. Точка, в которой происходит что-то неприятное: скажем, функция становится бесконечной или решение некоего уравнения прекращает существование.

Синус. Тригонометрическая функция угла, определяемая как $\sin A = b/c$ (см. рис. 51).

Скорость. Быстрота, с которой изменяется положение тела во времени. Скорость имеет как размер (абсолютную величину), так и направление.

Собственное число. Одно из нескольких особых чисел, связанных с оператором. Если при преобразовании некоего вектора при помощи этого оператора получается вектор, кратный первоначальному, то коэффициент кратности называется собственным числом.

Составное число. Натуральное число, которое можно получить перемножением двух меньших натуральных чисел.

Стандартная модель. Квантовомеханическая модель, описывающая все известные элементарные частицы.

Степенной ряд. То же, что многочлен, но с бесконечным количеством степеней переменной. К примеру, $1 + 2x + 3x^2 + + 4x^3 + \ldots$ В определенных обстоятельствах эта бесконечная сумма приобретает вполне определенное значение, и тогда говорят, что ряд сходится.

Степень. Число, умноженное само на себя заданное количество раз. К примеру, четвертая степень 3 — это $3 \times 3 \times 3 \times 3 = = 81$, обозначается 3^4.

Степень многочлена. Наибольшая степень переменной в многочлене. К примеру, степень многочлена $6x^3 – 5x^2 + 4x – 7$ равна 3.

Сфера. Множество всех точек в пространстве, расположенных на заданном расстоянии от некой фиксированной точки — центра. Она круглая, как мяч, но собственно сфера содержит только точки на поверхности мяча, а не внутри него.

Тангенс. Тригонометрическая функция угла, определяемая как $\operatorname{tg} A = b/a$ (см. рис. 51).

Топологическое пространство. Форма, которая считается «той же самой», если подвергается любому непрерывному преобразованию.

Топология. Наука о топологических пространствах.

Тор. Поверхность, похожая на бублик с одним отверстием (см. рис. 12).

Трансцендентное число. Число, не удовлетворяющее ни одному алгебраическому уравнению с рациональными коэффициентами. Примеры: π и e.

Трехмерная сфера. Трехмерный аналог сферы: множество всех точек четырехмерного пространства, лежащих на заданном расстоянии от некоей фиксированной точки — центра.

Триангуляция. Разбивка поверхности на сеть треугольников или его многомерный аналог.

Тривиальная группа. Группа, состоящая из единственного элемента, причем единичного.

Трисекция. Деление на три равных части, особенно в отношении углов.

Упаковка. Организация форм в пространстве таким образом, чтобы они не накладывались друг на друга.

Устойчивое состояние. Состояние динамической системы, в которое она возвращается, будучи подвергнута небольшому возмущению.

Фаза. Комплексное число на единичной окружности, на которое домножается квантовая волновая функция.

Фундаментальная группа. Группа, образованная гомотопическими классами петель в некоем топологическом пространстве с операцией «последовательное прохождение петель».

Функция. Правило f, которое при действии на число x дает другое число $f(x)$. К примеру, если $f(x) = \log x$, то f — логарифмическая функция. Переменная x может быть действительной или комплексной (в этом случае ее часто обозначают z). В более общем случае x и $f(x)$ могут быть элементами определенных множеств (в частности, плоскости или пространства).

Хаос. Случайное, на первый взгляд, поведение детерминированной системы.

Целое число. Любое из чисел … –3, –2, –1, 0, 1, 2, 3 …

Цикл. В топологии: формальная комбинация петель в триангуляции с присвоенными им числовыми индексами. В алгебраической геометрии: формальная комбинация подмногообразий с числовыми индексами.

Частица. Масса, сосредоточенная в одной точке.

Число Ферма. Число вида, где k — натуральное число. Если это число простое, оно называется *простым числом Ферма*.

Шар. Заполненная сфера, т. е. сфера и то, что находится у нее внутри.

Эйлерова характеристика. $F - E + V$, где F — число граней в триангуляции некоего пространства, E — число ребер, а V — число узлов. Для тора с g отверстиями эта величина равна $2 - 2g$ при любом разбиении на треугольники.

Электромагнитное поле. Функция, задающая силу и направление электрического и магнитного полей в каждой точке пространства.

Эллиптическая кривая. Кривая на плоскости, уравнение которой имеет вид $y^2 = ax^3 + bx^2 + cx + d$; постоянные a, b, c, d обычно считаются рациональными (см. рис. 27).

Эллиптическая функция. Комплексная функция, значение которой не меняется при прибавлении к переменной двух независимых комплексных чисел. Иными словами, $f(z) = f(z+u) = f(z+v)$, где v не равно u, домноженному на действительный коэффициент (см. рис. 30).

Примечания

[1] Вот как выглядят на сегодняшний день проблемы Гильберта и их статус:

1. **Континуум-гипотеза.** Существует ли бесконечное кардинальное число строго между кардиналами множеств целых и действительных чисел? Решена Полом Коэном в 1963 г. — ответ на вопрос зависит от того, какие аксиомы используются в теории множеств.
2. **Логическая непротиворечивость арифметики.** Доказать, что стандартные аксиомы арифметики не могут привести к противоречию. Решена Куртом Геделем в 1931 г.: с обычными аксиомами теории множеств такое доказательство невозможно.
3. **Равносоставленность равновеликих тетраэдров.** Если два тетраэдра имеют одинаковый объем, то всегда ли можно разрезать один из них на конечное число многоугольников и собрать из них второй? Решена в 1901 г. Максом Деном, ответ отрицательный.
4. **Прямая как кратчайшее расстояние между двумя точками.** Сформулировать аксиомы геометрии на основе данного определения прямой и посмотреть, что из этого следует. Слишком расплывчатая задача, чтобы можно было рассчитывать на определенное решение, но сделано немало.
5. **Группы Ли без опоры на дифференцируемость.** Технический вопрос теории групп преобразований. В одной из интерпретаций ее решил Эндрю Глисон в 1950-е гг., в другой — Хидехико Ямабе.
6. **Аксиомы физики.** Разработать строгую систему аксиом для математических областей физики, таких как теория вероятностей или механика. Систему аксиом для вероятностей построил Андрей Колмогоров в 1933 г.
7. **Иррациональные и трансцендентные числа.** Доказать, что определенные числа являются иррациональными или трансцендентными. Решена в 1934 г. Александром Гельфондом и Теодором Шнайдером.
8. **Гипотеза Римана.** Доказать, что все нетривиальные нули римановой дзета-функции лежат на критической линии. См. главу 9.
9. **Законы взаимности в числовых полях.** Обобщить классический закон квадратичной взаимности (о квадратах по определенному модулю) на более высокие степени. Частично решена.
10. **Условия существования решений диофантовых уравнений.** Найти алгоритм, позволяющий определить, имеет ли данное

полиномиальное уравнение со многими переменными решения в целых числах. Невозможность доказал Юрий Матиясевич в 1970 г.

11. **Квадратичные формы с алгебраическими числами в качестве коэффициентов.** Технические вопросы решения диофантовых уравнений со многими переменными. Решена частично.
12. **Теорема Кронекера об абелевых полях.** Технические вопросы обобщения теоремы Кронекера. Не доказана до сих пор.
13. **Решение уравнений седьмой степени при помощи функций специального вида.** Доказать, что общее уравнение седьмой степени не может быть решено с использованием функций двух переменных. В одной из интерпретаций возможность такого решения доказали Андрей Колмогоров и Владимир Арнольд.
14. **Конечность полной системы функций.** Расширить теорему Гильберта об алгебраических инвариантах на все группы преобразований. Опроверг Масаёси Нагата в 1959 г.
15. **Исчислительная геометрия Шуберта.** Герман Шуберт нашел нестрогий метод подчета различных геометрических конфигураций. Задача в том, чтобы сделать этот метод строгим. Полного решения до сих пор нет.
16. **Топология кривых и поверхностей.** Сколько связанных компонент может иметь алгебраическая кривая заданной степени? Сколько различных периодических циклов может иметь алгебраическое дифференциальное уравнение заданной степени? Ограниченное продвижение.
17. **Представление определенных форм в виде суммы квадратов.** Если рациональная функция всегда принимает неотрицательные значения, то должна ли она обязательно выражаться в виде суммы квадратов? Решили Эмиль Артин, Д. Дюбуа и Альбрехт Пфистер. Верно для действительных чисел, неверно в некоторых других числовых системах.
18. **Заполнение пространства многогранниками.** Общие вопросы о заполнении пространства конгруэнтными многогранниками. Имеет отношение к гипотезе Кеплера, ныне доказанной (см. главу 5).
19. **Аналитичность решений в вариационном исчислении.** Вариационное исчисление отвечает на такие вопросы, как «найти кратчайшую кривую с заданными свойствами». Если подобная задача формулируется при помощи красивых функций, то должно ли решение тоже быть красивым? Доказали Эннио де Джорджи в 1957 г. и Джон Нэш.
20. **Граничные задачи.** Разобраться в решениях дифференциальных уравнений физики в определенной области пространства, если заданы свойства решения на ограничивающей эту область поверхности. В основном решена (вклад внесли многие математики).

21. **Существование дифференциальных уравнений с заданной монодромией.** Особый тип комплексного дифференциального уравнения, в котором можно разобраться при помощи данных о его точках сингулярности и группе монодромии. Доказать, что может существовать любая комбинация этих данных. Ответ «да» или «нет» в зависимости от интерпретации.
22. **Униформизация с использованием автоморфных функций.** Технический вопрос об упрощении уравнений. Решил Пауль Кебе вскоре после 1900 г.
23. **Развитие вариационного исчисления.** Гильберт призывал к выдвижению новых идей в области вариационного исчислении. Многое сделано, но формулировка слишком неопределенная, чтобы задачу можно было считать решенной.

[2] Алгоритм Агравала–Каяла–Саксены выглядит так:

- Если n представляет собой точную степень меньшего числа, выдаем СОСТАВНОЕ.
- Находим наименьшее r, такое, что наименьшая степень r, равная 1 по модулю n, больше или равна $(\log n)^2$.
- Если какое-либо число, меньшее или равное r, имеет общий делитель с n, выдаем СОСТАВНОЕ.
- Если n меньше или равно r, выдаем ПРОСТОЕ.
- Для всех целых чисел a от 1 до определенного предела проверяем, совпадает ли многочлен $(x + a)^n$ с многочленом $x^n + a$ по модулю n и по модулю $x^r - 1$. Если в обоих случаях ответ положительный, выдаем СОСТАВНОЕ.
- Выдаем ПРОСТОЕ.

[3] Примером того, что я имею в виду, может служить формула, где квадратные скобки обозначают наибольшее целое число, меньшее или равное их содержимому. В 1947 г. У. Миллс доказал, что существует действительная константа A, такая, что для любого n вычисленное по этой формуле значение будет простым. Если считать гипотезу Римана верной, то минимальное значение A, удовлетворяющее условию, равно приблизительно 1,306. Однако эта константа определяется при помощи подходящей последовательности простых чисел, а формула — всего лишь символьный способ записи этой последовательно-

сти. Подобные формулы, включая некоторые из тех, что представляют все простые числа, представлены также на сайтах:

http://mathworld.wolfram.com/PrimeFormulas.html;
http://en.wikipedia.org/wiki/Formula_for_primes.

[4] Если n — нечетное, то $n - 3$ четное, а если n больше 5, то $n - 3$ больше 2. Согласно первой гипотезе, $n - 3 = p + q$, так что $n = p + q + 3$.

[5] Можно упомянуть в этом контексте выражение «квантовый скачок». В повседневной речи оно обычно означает какой-то гигантский шаг вперед или резкую перемену, как, например, открытие европейцами Америки. Однако в квантовой теории квантовый скачок настолько мал, что ни один известный инструмент не способен зарегистрировать его непосредственно; изменение при этом выражается 0,000…01 примерно с 40 нулями.

[6] Нахождение конечного разбиения квадрата и собирания из этих частей круга известно как квадратура круга Тарского. Миклош Лацкович решил эту задачу в 1990 г. Его метод неконструктивен и использует теорему выбора, при этом число частей, на которые нужно делить квадрат, огромно — около 10^{50}.

[7] Квадратриса Гиппия — это кривая, описываемая точкой пересечения вертикальной прямой, движущейся равномерно через прямоугольник, и прямой, которая равномерно поворачивается вокруг середины нижней стороны прямоугольника (см. рис. 52). Такое соотношение превращает любой вопрос о делении угла в вопрос о соответствующем делении отрезка. К примеру, чтобы разделить угол натрое, нужно всего лишь разделить натрое соответствующий отрезок прямой. См.: http://www.geom.uiuc.edu/~huberty/math5337/groupe/quadratrix.html.

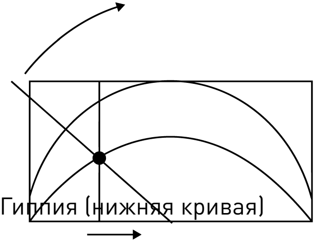
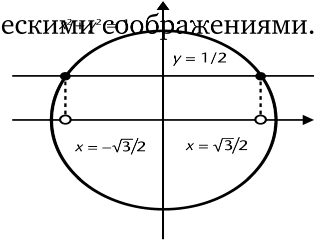


Рис. 52. Квадратриса Гиппия (нижняя кривая)

[8] Вот красноречивый пример. Геометрически если прямая пересекается с окружностью и не является касательной, то она имеет с окружностью ровно две общие точки. Возьмем прямую, параллельную горизонтальной оси, на расстоянии 1/2 над ней (см. рис. 53). Эта прямая описывается очень простым уравнением: $y = 1/2$. (При любом x мы имеем одно и то же значение y.) Если $y = 1/2$, то уравнение $x^2 + y^2 = 1$ превращается в $x^2 + 1/4 = 1$. Отсюда $x^2 = 3/4$, а $x = \frac{3}{2}$ или $-\frac{3}{2}$. Алгебра говорит, что прямая пересекает единичную окружность ровно в двух точках $\left(\frac{3}{2}, \frac{1}{2}\right)$ и $-\left(\frac{3}{2}, \frac{1}{2}\right)$. Это вполне согласуется с рис. 53 и чисто геометрическими соображениями.



[9] Строго говоря, многочлен, о котором идет речь, должен иметь целые коэффициенты и быть несокращаемым (т. е. не являться произведением двух многочленов меньших степеней с целыми коэффициентами). Степень многочлена, равная степени двойки, — необходимое, но не достаточное условие для существования построения при помощи циркуля и линейки. Если степень не равна степени двойки, построение существовать не может. Если равна, то для решения вопроса о его существовании необходим дальнейший анализ.

[10] Обратное тоже верно: данные построения для правильных трех- и пятиугольников можно получить из построения 15-угольника. Идея в том, что $2/5 - 1/3 = 1/15$. В отношении простых степеней есть один тонкий момент. Эти рассуждения не позволяют построить, скажем, девятиугольник, хотя построение для простых делителей числа (а именно треугольника) существует. Гаусс доказал, что для нечетных простых чисел, возведенных в степень больше 1, построение невозможно.

[11] Чтобы разобраться в этом утверждении, разложим квадратный многочлен на линейные множители. Тогда $x^2 - 1 = (x + 1)(x - 1)$, что равно нулю, если любой из множителей равен нулю, так что $x = 1$ или $x = -1$. Те же рассуждения можно применить к $x^2 = xx$: это равно нулю, если нулю равен один из множителей. В данном случае они совпадают, но наличие двух множителей x отличает этот случай от чего-нибудь вроде $x\,(x - 1)$, где множитель x один. При ответе на вопрос о том, сколько решений имеет алгебраическое уравнение, подобную «множественность» лучше учитывать.

[12] При $n = 9$ второй множитель будет

$$x^8 + x^7 + x^6 + x^5 + x^4 + x^3 + x^2 + x + 1.$$

Но он и сам является составным: он равен

$$(x^2 + x + 1)\,(x^6 + x^3 + 1).$$

Гауссова характеристика чисел, допускающих построение, требует, чтобы степень каждого несократимого множителя была степенью 2. Но степень второго множителя — 6 — не является степенью 2.

[13] Гаусс доказал, что 17-угольник можно построить, если вы умеете строить отрезки длиной

$$\frac{1}{16}\Big[-1+\quad 17+\quad 34-2\quad 17+$$

$$+\quad 68+12\quad 17-16\quad 34+2\quad 17-2(1-\quad 17)(\quad 34-2\quad 17)\Big].$$

Поскольку квадратный корень всегда можно построить, это вполне эффективно решает задачу. Другие математики нашли более очевидные построения. Ульрих фон Гугенин опубликовал первое из них в 1803 г., а Г. Ричмонд в 1893 г. нашел более простое. На рис. 54 возьмем два перпендикулярных радиуса AOP_0 и BOC окружности. Пусть OJ = 1/4 OB, а угол OJE = 1/4 OJP_0. Найдем F, такое, что угол EJF равен 45°. Построим окружность с диаметром FP_0; она пересекается с OB в точке K. Проведем через K окружность с центром в точке E; она пересечет AP_0 в точках G и H. Построим в этих точках перпендикуляры к AP_0, назовем их HP_3 и GP_5. Тогда P_0, P_3, P_5 представляют собой соответственно нулевую, третью и пятую вершины правильного 17-угольника. Теперь несложно построить и остальные вершины.

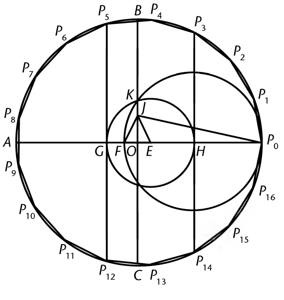


Рис. 54. Как построить правильный 17-угольник

[14] Ф. Ришло опубликовал построение правильного 257-угольника в 1832 г. Йоханн Хермес посвятил десять лет жизни исследованию 65 537-угольника. Его неопубликованную работу можно найти в Университете Геттингена, но считается, что в ней есть ошибки.

[15] Типичная цепная дробь выглядит примерно так:

$$3+\cfrac{1}{7+\cfrac{1}{15+\cfrac{1}{1+\cfrac{1}{292+\dots}}}}$$

Данная конкретная цепная дробь представляет собой начало дроби, представляющей число π.

[16] Если границы могут быть по-настоящему сложными, не как на карте, а гораздо более извилистыми, то общую «границу» могут иметь сколько угодно стран. Этот неочевидный результат иллюстрирует конструкция, известная как «озера Вады».

[17] До недавнего времени статья в *Nature* считалась последней публикацией, посвященной этой проблеме, почти за 100 лет, но историк математики Робин Уилсон отыскал более позднюю статью Кейли.

[18] При работе в двойственной сети пусть F — число граней (включая одну большую грань, окружающую сеть целиком), E — число ребер, а V — число вершин. Можно считать, что каждая грань двойственной сети имеет по крайней мере три ребра — ведь если в ней есть грань только с двумя ребрами, то она соответствует «лишней» вершине первоначальной сети, в которой встречаются всего два ребра. Такую вершину можно удалить, а два ребра объединить в одно.

Каждое ребро граничит с двумя гранями, и каждая грань имеет по крайней мере три ребра, потому $E \geq 3F/2$ или, что то же самое, $2E/3 \geq F$. Согласно уравнению Эйлера, $F + V - E = 2$, так что $2E/3 + V - E \geq 2$. Из этого следует, что $12 + 2E \leq 6V$.

Пусть V_m — это число вершин с m соседями. Тогда $V = V_6 + V_7 + V_8 + \ldots$

Поскольку каждое ребро соединяет две вершины:

$$2E = 6V_6 + 7V_7 + 8V_6 + \ldots$$

Подставив в неравенство, получаем:

$$12 + 6V_6 + 7V_7 + 8V_8 + \ldots \leq 6V_6 + 6V_7 + 6V_8 + \ldots,$$

так что $12 + V_7 + 2V_8 + \ldots \leq 0$,
что невозможно.

[19] Термин «цепочка» здесь неточен, потому что предполагает линейную последовательность. Цепочка Кемпе может содержать петли и разветвляться.

[20] Поскольку пространство бесконечно, шариков в нем тоже бесконечно много, так что общий объем и пространства, и шариков также бесконечен. Мы не можем определить плотность как ∞/∞, потому что эта величина математически не определена. Вместо этого мы последовательно рассматриваем все более крупные области пространства и берем верхний предел отношения областей, заполненных шариками.

[21] Хейлс использовал несколько различных определений того, что я называю клеткой. Последнее — «звезда декомпозиции». В моем описании опущены некоторые принципиально важные отличия; так общая идея получается более понятной.

[22] Пусть область представляет собой многоугольник, как на рис. 55. Для любой точки, не лежащей на линиях многоугольника, существует проходящая через нее прямая, которая выходит за пределы описывающей многоугольник окружности и не проходит ни через одну его вершину. (Вершин — конечное количество, а прямых — бесконечное, есть из чего выбрать.) Эта прямая пересекает многоугольник конечное число раз, причем число это либо четное, либо нечетное. Определим, что внутренняя часть состоит из точек, для которых это число нечетное, а внешняя — из точек, для которых оно четное. Без труда доказывается, что каждая из этих областей является связной, а многоугольник их разделяет (см. рис. 55).

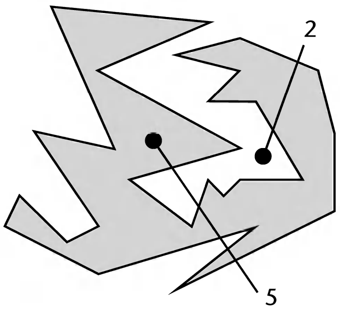


Рис. 55. Доказательство теоремы Жордана о кривых. Нечетным число пересечений является для точек в серой области (*внутри*), а четным — для точек в белой области (*снаружи*)

[23] Расширим это загадочное замечание: вот формула

$$\int \frac{\mathrm{d}x}{1-x^2} = \arcsin x,$$

где арксинус (arcsin или $\sin^{-1}$) представляет собой функцию, обратную синусу. Иными словами, если $y = \sin x$, то $x = \arcsin y$.

[24] К примеру, пусть k — любое комплексное число. Рассмотрим интеграл

$$\int \frac{\mathrm{d}x}{(1-x^2)(1-k^2x^2)}.$$

Это функция, обратная эллиптической функции, обозначаемой как sn. Такая функция существует для каждого значения k. Устроено все примерно так же, как в случае с sin и arcsin, но хитрее.

[25] Один корень p-й степени из единицы равен комплексному числу

$$\zeta = \cos 2\pi/p + \mathrm{i}\sin 2\pi/p,$$

а остальные представляют собой его степени $\zeta^2, \zeta^3, \ldots \zeta^{p-1}$. Чтобы понять почему, вспомните, что тригонометрические функции синус и косинус определяются через прямоугольный треугольник (см. рис. 56 слева). Обозначив стороны треугольника традиционными a, b, c, мы определяем синус и косинус угла A как $\sin A = a/c$, $\cos A = b/c$.

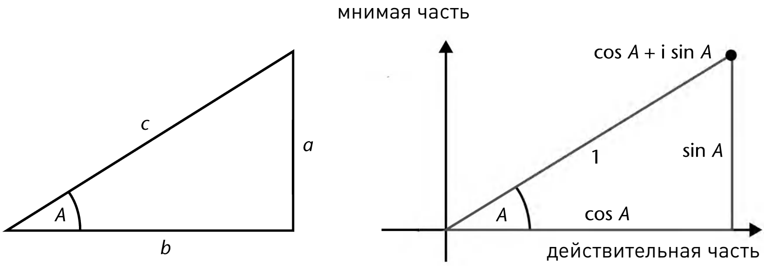


Рис. 56. Определение синуса и косинуса *(слева)*. Интерпретация на комплексной плоскости *(справа)*

Если мы возьмем $c = 1$ и поместим этот треугольник на комплексную плоскость, как на рис. 56 справа, то вершина, в которой встречаются c и a, представляет собой точку

$$\cos A + \mathrm{i} \sin A.$$

Несложно доказать, что для любых углов A и B

$$(\cos A + \mathrm{i} \sin A)\ (\cos B + \mathrm{i} \sin B) = \cos\ (A + B) + \mathrm{i} \sin\ (A + B),$$

а это ведет непосредственно к формуле Муавра

$$(\cos A + \mathrm{i} \sin A)^n = (\cos nA + \mathrm{i} \sin nA)$$

для любого натурального n. Поэтому

$$\zeta^p = (\cos 2\pi/p + \mathrm{i} \sin 2\pi/p)^p = \cos 2\pi + \mathrm{i} \sin 2\pi = 1$$

для любой степени 1, где $\zeta, \zeta^2, \zeta^3, \ldots, \zeta^{p-1}$ есть корень p-й степени из единицы. На этом мы остановимся, поскольку $\zeta^p = 1$ и, соответственно, для более высоких степеней новых чисел не появится.

[26] Введем понятие *нормы.*

$$\mathrm{N}(a + b \quad 15\ ) = a^2 - 15b^2,$$

имеющее замечательное свойство

$$\mathrm{N}(xy) = \mathrm{N}(x)\mathrm{N}(y).$$

Тогда

$$\mathrm{N}(2) = 4\mathrm{N}(5) = 25\mathrm{N}(5 + \quad 15\ ) = 10\mathrm{N}(5 - \quad 15\ ) = 10.$$

Любой собственный делитель любого из этих четырех чисел должен иметь норму 2 или 5 (собственные делители их норм). Но уравнения $a^2 - 15b^2 = 2$ и $a^2 - 15b^2 = 5$ не имеют целых решений. Следовательно, собственных делителей не существует.

[27] А может быть, и нет. Владимир Кривченков установил, что энергия основного состояния и первых возбужденных состояний для квантовой задачи трех тел может быть рассчитана вручную. Но в классической механике аналогичная проблема оказывается менее решаемой из-за хаоса.

[28] Более формально это называется временем Ляпунова.

[29] Существует вариант, где $1/\log t$ интегрируется от 2 до x, а не от 0 до x. Это помогает обойти технически сложный момент при $t = 0$, где $\log t$ не определен. Иногда для этого варианта используется обозначение $\mathrm{Li}(x)$, а определенная в тексте функция называется $\mathrm{Li}(x)$.

[30] Это следует из занятной формулы Римана

$$\zeta(1-s)=2^{1-s}\pi^{-s}\sin\left(\frac{\pi(1-s)}{2}\right)\Gamma(s)\zeta(s),$$

где Γ (s) — классическая функция, известная как гамма-функция и определенная для всех комплексных s. Правая ее сторона определена, когда действительная часть s больше 1.

[31] Риман определил еще одну похожую функцию

$$\Pi(x)=\pi(x)+\frac{1}{2}\pi(x^{1/2})+\frac{1}{3}\pi(x^{1/3})+\frac{1}{4}\pi(x^{1/4})+\ldots .$$

которая подсчитывает скорее простые степени, чем простые числа. Из этого можно восстановить $\pi(x)$. Затем он доказал точную формулу для этой модифицированной функции в терминах логарифмических интегралов и связанного интеграла:

$$\Pi(x)=\mathrm{Li}(x)-\sum_{\rho}\mathrm{Li}(x^{\rho})+\int_{x}^{\infty}\frac{\mathrm{d}t}{t(t^{2}-1)\log t}$$

Здесь Σ — сумма по всем значениям ρ, при которых ζ (ρ) равна 0, исключая отрицательные целые числа.

[32] К примеру, асимптотична $x+\quad x$: их отношение равно

$$(x+\quad x)/\mathrm{x}=1+1/\quad x.$$

С ростом x растет и $\quad x$, так что $1/\quad x$ стремится к 0, а отношение стремится к 1. Но разность составляет $\quad x$ и становится все больше с ростом x. К примеру, когда x равен 1 трлн, $\quad x$ равен 1 млн.

[33] Постоянная Эйлера — это предел при n, стремящемся к бесконечности выражения

$$1+\frac{1}{2}+\frac{1}{3}+\ldots+\frac{1}{n}-\log n.$$

[34] Единичная трехмерная сфера содержит все точки с координатами (x, y, z, w), такими, что $x^2 + y^2 + z^2 + w^2 = 1$. Есть несколько способов сделать трехмерную сферу более интуитивно понятной. Ее можно рассматривать по аналогии с двумерной сферой и проверять посредством координатной геометрии. Одно такое описание («заполненный шар, все точки поверхности которого тождественны между собой») дано в тексте, а на рис. 57 можно увидеть еще одно. Чтобы установить аналогию, заметим, что при разрезании двумерной сферы вдоль экватора мы получаем две полусферы. Каждая из них непрерывно преобразуется в плоский диск. Чтобы восстановить двумерную сферу, мы просто отождествляем соответствующие точки на границах этих двух дисков. В каком-то смысле мы изготовили карту двумерной сферы из двух плоских дисков, так же как картографы создают плоские проекции нашей круглой планеты. Трехмерную сферу можно построить при помощи аналогичной процедуры. Возьмем два заполненных шара и отождествим между собой соответствующие точки их поверхностей. Теперь оба они имеют одну и ту же поверхность (ведь мы их отождествили), и эта поверхность представляет собой двумерную сферу. Она образует «экватор» трехмерной сферы.

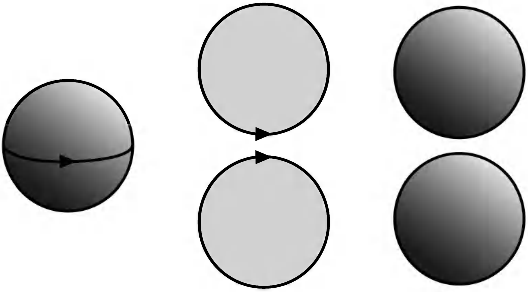


Рис. 57. Как изготовить трехмерную сферу: *слева* — разрезаем двумерную сферу на полусферы; *в середине* — восстанавливаем двумерную сферу из двух половинок путем склеивания краев; *справа* — по аналогии концептуально склеиваем поверхности двух шаров воедино, так что соответствующие точки считаются идентичными. Получается трехмерная сфера

[35] Традиционно считается, что мы говорим о сложении и используем обозначение $a + b$, когда коммутативный закон соблюдается, но говорим об умножении и используем обозначение ab, когда он может и не соблюдаться. Здесь, однако, я не следовал этому принципу, поскольку эта книга — не учебник по теории групп, а термин «сложение» представляется более естественным.

[36] Начнем счет с нуля. Всякий раз, проходя остановку на пути против часовой стрелки, увеличиваем счет на 1; проходя ее против часовой стрелки, уменьшаем на 1. В конце поездки добавляем 1, если приехали против часовой стрелки, и вычитаем 1, если приехали по часовой стрелке. Окончательный результат подсчета скажет вам, сколько раз вы обогнули окружность в общем направлении против часовой стрелки.

[37] Формула Стирлинга утверждает, что $n!$ приблизительно равно

$$2\pi n \left(n / e\right)^n .$$

[38] Леонардо нашел семейство решений

$$\left(\frac{m^2+n^2}{2}\right)^2 - mn\left(m^2-n^2\right)^2 = \left(\frac{m^2-2mn-n^2}{2}\right)^2$$
$$\left(\frac{m^2+n^2}{2}\right)^2 + mn\left(m^2-n^2\right)^2 = \left(\frac{m^2+2mn-n^2}{2}\right)^2,$$

где m, n — нечетные числа. Роль d здесь выполняется числом $mn\ (m^2 - n^2)$, а x равен $m^2 + n^2/2$. Выбор $m = 5$, $n = 4$ ведет к тому, что $mn\ (m^2 - n^2) = 720$. Более того, $720 = 5 \times 12^2$. Разделив x на 12, получим ответ.

[39] Если $x - n$, x и $x + n$ представляют собой квадраты, то их произведение тоже квадрат и равно $x^3 - n^2x$. Следовательно, уравнение $y^2 = x^3 - n^2x$ имеет рациональное решение. Более того, y не равен нулю, в противном случае $x = n$, т. е. x и $2x$ — квадраты, что невозможно, поскольку число 2 иррационально.

Напротив, если x и y удовлетворяют кубическому уравнению и y не равен 0, то $a = (x^2 - n^2)/y$, $b = 2nx/y$ и $c = (x^2 + n^2)/y$ удовлетворяют уравнениям $a^2 + b^2 = c^2$ и $ab/2 = n$.

[40] Иными словами:

$$\prod_{p \leq x} \frac{N_p}{p} \approx C(\log r)^r,$$

где r — ранг, C — константа, а знак $\approx$ означает, что отношение двух частей этого выражения стремится к 1 при x, стремящемся к бесконечности.

[41] Наиболее вероятная причина заключается в том, что существуют естественные варианты перевода с соответствующих языков, которыми пользуются лучшие математики обеих областей.

[42] Я не знаю, почему b — не число бананов. Возможно, потому, что в послевоенной Британии бананы были экзотикой и в продаже попадались редко.

[43] Отсюда известная математическая шутка. Биолог, статистик и математик сидели на открытой площадке кафе и наблюдали за тем, что происходит вокруг. Вот в дом напротив зашли мужчина и женщина, а через 10 минут вышли снова уже с ребенком. «Они размножились», — заметил биолог. «Нет, — возразил статистик. — Это ошибка наблюдения. В среднем и туда, и обратно прошли два с половиной человека». «Нет, нет, нет, — сказал математик. — Все совершенно очевидно. Если теперь в здание кто-то войдет, оно станет пустым».

[44] Может быть, Бор здесь попал в точку. Научные теории проверяются через предсказания, но мало какие из них способны предсказать будущее. Большинство ограничивается утверждениями вида «если, то»: если пропустить свет сквозь призму, то он расщепится на отдельные цвета. Это «предсказание» ничего не говорит о том, когда это произойдет. Поэтому, как ни парадоксально, мы можем предсказывать погоду, не предсказывая ее. «Если теплый воздух циклона встретится с холодным воздухом, пойдет снег». Это, конечно, научное предсказание, но отнюдь не прогноз погоды.

[45] Гипотеза ABC гласит: для любого $\varepsilon > 0$ существует постоянная $k\varepsilon>0$, такая, что если a, b и c — положительные целые числа, не имеющие общих делителей больших 1, и $a + b = c$, то $c \leq k\varepsilon P^{1+}\varepsilon$, где P — результат перемножения всех различных простых делителей abc.

[46] В сентябре 2012 г. Синъити Мотидзуки объявил, что ему удалось доказать гипотезу ABC при помощи кардинально нового подхода к основным положениям алгебраической геометрии. Специалисты проверяют его 500-страничное доказательство, но это может занять довольно много времени.

Предметный указатель

Стюарт Иэн

Величайшие математические задачи

Руководитель проекта *И. Серёгина*
Корректоры *Е. Аксёнова, М. Миловидова*
Компьютерная верстка *А. Фоминов*
Дизайн обложки *О. Сидоренко*

Подписано в печать 24.10.2014. Формат 60×90/16.
Бумага офсетная № 1. Печать офсетная.
Объем 29 печ. л. Тираж 5000 экз. Заказ №   .

ООО «Альпина нон-фикшн»
123060, г. Москва
ул. Расплетина, д. 19, офис 2
Тел. (495) 980-5354
www.nonfiction.ru

Знак информационной продукции
(Федеральный закон № 436-ФЗ от 29.12.2010 г.) 0+

ИЭН СТЮАРТ — профессор Математического института Уорикского университета, известный популяризатор науки, автор 80 книг и множества научно-популярных статей. Его научные интересы лежат в области теории катастроф, симметрии, теории групп и теории бифуркаций.

На протяжении тысячелетий в математике существовали задачи, работа над которыми открывала совершенно новые направления и возможности. Именно поэтому над их решением бились величайшие умы, и тем не менее некоторые такие задачи до сих пор остаются нерешенными. Английский математик Иэн Стюарт рассказывает об истоках этих математических проблем; о том, почему они так важны и какое место занимают в общем контексте математики и естественных наук.

Вы узнаете об Эндрю Уайлсе и его семилетних поисках разгадки теоремы Ферма, которая оставалась неприступной 350 лет; об эксцентричном гении Григории Перельмане, отказавшемся от приза в миллион долларов за доказательство гипотезы Пуанкаре, а также о других блестящих математиках современности, которые принимают вызовы своих предшественников и создают науку будущего.

> Стюарт блестяще разъясняет самые сложные проблемы.

New Scientist

> Образные, остроумные истории Стюарта, аналогии и диаграммы проясняют чрезвычайно сложные понятия. Это приведет в восторг как энтузиастов математической науки, так и широкий круг читателей, которым интересна эта тема.

Kirkus Reviews

> Описанный Стюартом удивительный мир математической науки и ведущихся в ней исследований выглядит поистине захватывающе и даже романтично.

The Times

> Он умеет сделать уравнения интересными, а мир науки — исключительно увлекательным.

New York Journal of Books

ЭЛЕМЕНТЫ
elementy.ru
Династия